



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial

(21) PI 0714870-4 A2



* B R P I 0 7 1 4 8 7 0 A 2 *

(22) Data de Depósito: 20/07/2007

(43) Data da Publicação: 28/05/2013
(RPI 2212)

(51) Int.Cl.:

C07K 14/39

C12N 9/14

C12N 15/80

(54) Título: MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO SECRETADO TENDO ATIVIDADE BIOLÓGICA, PROTEÍNA DE FUSÃO ISOLADA, POLINUCLEOTÍDEO ISOLADO, CONSTRUÇÃO DE PROTEÍNA DE FUSÃO, VETOR DE EXPRESSÃO, CÉLULA HOSPEDEIRA FÚNGICA, MÉTODOS PARA DEGRADAR OU CONVERTER UM MATERIAL CELULÓSICO E PARA PRODUZIR UMA SUBSTÂNCIA

(30) Prioridade Unionista: 21/07/2006 US 60/832511

(73) Titular(es): Novozymes, Inc.

(72) Inventor(es): Sandra Merino

(74) Procurador(es): Momsen, Leonardos & Cia

(86) Pedido Internacional: PCT US2007074038 de 20/07/2007

(87) Publicação Internacional: WO 2008/057637de 15/05/2008

(57) Resumo: MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO SECRETADO TENDO ATIVIDADE BIOLÓGICA, PROTEÍNA DE FUSÃO ISOLADA, POLINUCLEOTÍDEO ISOLADO, CONSTRUÇÃO DE PROTEÍNA DE FUSÃO, VETOR DE EXPRESSÃO, CÉLULA HOSPEDEIRA FÚNGICA, MÉTODOS PARA DEGRADAR OU CONVERTER UM MATERIAL CELULÓSICO E PARA PRODUZIR UMA SUBSTÂNCIA. A presente invenção refere-se a métodos para produzir um polipeptídeo secretado tendo atividade biológica, compreendendo: (a) transformar uma célula hospedeira fúngica com uma construção de proteína de fusão codificando uma proteína de fusão, que compreende: (i) um primeiro polinucleotídeo codificando um peptídeo de sinal; (ii) um segundo polinucleotídeo codificando pelo menos um domínio catalítico de uma endoglucanase ou uma porção do mesmo; e (iii) um terceiro polinucleotídeo codificando pelo menos um domínio catalítico de um polipeptídeo tendo atividade biológica; em que o peptídeo de sinal e pelo menos o domínio catalítico da endoglucanase aumenta a secreção do polipeptídeo tendo atividade biológica comparado com a ausência de pelo menos o domínio catalítico da endoglucanase; (b) cultivar a célula hospedeira fúngica transformada sob condições apropriadas para produção da proteína de fusão; e (c) recuperar a proteína de fusão, um componente da mesma, ou uma combinação da mesma; tendo atividade biológica, a partir do meio de cultivo.

“MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO SECRETADO TENDO ATIVIDADE BIOLÓGICA, PROTEÍNA DE FUSÃO ISOLADA, POLINUCLEOTÍDEO ISOLADO, CONSTRUÇÃO DE PROTEÍNA DE FUSÃO, VETOR DE EXPRESSÃO, CÉLULA HOSPEDEIRA FÚNGICA, MÉTODOS PARA DEGRADAR OU CONVERTER UM MATERIAL CELULÓSICO E PARA PRODUZIR UMA SUBSTÂNCIA”

Declaração com Relação aos Direitos às Invenções Realizadas sob Pesquisa e Desenvolvimento Financiados pelo Governo Federal

Esta invenção foi feita com apoio governamental sob NREL Subcontrato no. ZCO-30017-02, Contrato Principal DE-AC36-98GO10337, concedido pelo Departamento de Energia. O Governo Americano tem determinados direitos sobre esta invenção.

Antecedentes da Invenção

Campo da Invenção

A presente invenção refere-se a métodos de produzir polipeptídeos secretados tendo atividade biológica e a proteínas de fusão e polinucleotídeos dos mesmos.

Descrição da Arte Relacionada

A produção recombinante de um polipeptídeo heterólogo em uma célula hospedeira fúngica, particularmente uma célula fúngica filamentosa como *Aspergillus* ou *Trichoderma* uma célula de levedura como *Saccharomyces*, pode prover um veículo mais desejável para produzir o polipeptídeo em quantidades comercialmente relevantes.

A produção recombinante de um polipeptídeo heterólogo secretado é geralmente obtida por construção de um cassete de expressão em que o DNA codificando para o polipeptídeo é ligado operativamente a um promotor apropriado para a célula hospedeira e uma região de codificação de peptídeo sinal que codifica para uma sequência de aminoácido ligada em quadro ao término amino do polipeptídeo e dirige o polipeptídeo codificado

na via secretória das células. O cassete de expressão é introduzido na célula hospedeira, geralmente por transformação mediada por plasmídeo. A produção da proteína heteróloga secretada é então obtida por cultivo da célula hospedeira transformada sob condições indutíveis necessárias para o funcionamento correto do promotor contido sobre o cassete de expressão.

Apesar da expressão de um polipeptídeo heterólogo em uma célula hospedeira poder ser melhorada, um obstáculo com frequência encontrado é que o polipeptídeo é fracamente secretado no meio de cultura. Um método de melhorar a secreção do polipeptídeo é substituir a sequência de codificação de peptídeo sinal nativa com uma região de codificação de peptídeo sinal estranho para melhorar a secreção do polipeptídeo. No entanto, em alguns casos, esta substituição não provê uma melhora suficiente para produzir o polipeptídeo em quantidades relevantes. Outro método consiste em fundir o polipeptídeo em outro polipeptídeo que é altamente secretado por uma célula hospedeira. O polipeptídeo altamente secretado funciona como um veículo para transportar o polipeptídeo fracamente secretado ou não secretado como uma proteína de fusão através da via secretória das células.

WO 05/093050 descreve uma proteína de fusão composta de um domínio catalítico exo-celobiohidrolase e um domínio catalítico de celulase para aumentar o rendimento de uma enzima celulase. Gouka et al, 1997, *Applied and Environmental Microbiology* Fev. 1997, p 488-497, descreve fusões de gene glucoamilase que aliviam as limitações para a produção de proteínas em *Aspergillus awamori*. Nyssonen e Keranen, 1995 *Current Genetics* 10 28 71-79, descreve papéis múltiplos de celobiohidrolase I na melhora da produção de anticorpos de fusão por *Trichoderma reesei*.

É um objeto da presente invenção prover métodos para aumentar a secreção de polipeptídeos tendo atividade biológica.

Sumário da Invenção

A presente invenção refere-se a métodos para produzir um

polipeptídeo secretado tendo atividade biológica, compreendendo

(a) transformar uma célula hospedeira fúngica com uma construção de proteína de fusão codificando uma proteína de fusão, em que a construção de proteína de fusão compreende

5 (i) um primeiro polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal,

(ii) um segundo polinucleotídeo compreendendo uma sequência de nucleotídeos codificando pelo menos um domínio catalítico de uma endoglucanase ou uma porção da mesma, e

10 (iii) um terceiro polinucleotídeo compreendendo uma sequência de nucleotídeos codificando pelo menos um domínio catalítico de um polipeptídeo tendo atividade biológica ou uma porção do mesmo, em que o peptídeo sinal e pelo menos o domínio catalítico da endoglucanase ou a porção do mesmo aumenta secreção do polipeptídeo tendo atividade biológica
15 ou a porção do mesmo comparado com a ausência de pelo menos o domínio catalítico da endoglucanase ou a porção do mesmo,

(b) cultivar a célula hospedeira fúngica transformada sob condições apropriadas para a produção da proteína de fusão, e

(c) recuperar a proteína de fusão, um componente da mesma,
20 ou uma combinação da mesma, a partir do meio de cultivo, em que a proteína de fusão ou o componente da mesma tem atividade biológica.

Em um aspecto, a extremidade 3' do primeiro polinucleotídeo é ligada operativamente à extremidade 5' do segundo polinucleotídeo e a extremidade 3' do segundo polinucleotídeo é ligada operativamente à
25 extremidade 5' do terceiro polinucleotídeo ou a extremidade 3' do primeiro polinucleotídeo é ligada operativamente à extremidade 5' do terceiro polinucleotídeo e a extremidade 3' do terceiro polinucleotídeo é ligada operativamente à extremidade 5' do segundo polinucleotídeo para codificar uma proteína de fusão.

A presente invenção também refere-se a proteínas de fusão isoladas, compreendendo:

(a) uma primeira sequência de aminoácidos compreendendo um peptídeo sinal,

5 (b) uma segunda sequência de aminoácidos compreendendo pelo menos um domínio catalítico de uma endoglucanase ou uma porção da mesma, e

(c) uma terceira sequência de aminoácidos compreendendo pelo menos um domínio catalítico de um polipeptídeo tendo atividade
10 biológica ou uma porção do mesmo.

A presente invenção também refere-se a polinucleotídeos codificando as proteínas de fusão, e construções de proteína de fusão, vetores de expressão e células hospedeiras recombinantes compreendendo tais polinucleotídeos.

15 Em outro aspecto, a extremidade C-terminal da primeira sequência de aminoácidos é ligada em armação à extremidade N-terminal da segunda sequência de aminoácidos e a extremidade C-terminal da segunda sequência de aminoácidos é ligada em armação à extremidade N-terminal da
20 terceira sequência de aminoácidos ou a extremidade C-terminal da primeira sequência de aminoácidos é ligada em armação à extremidade N-terminal da terceira sequência de aminoácidos e a extremidade C-terminal da terceira sequência de aminoácidos é ligada em armação à extremidade N-terminal da segunda sequência de aminoácidos.

25 A presente invenção ainda refere-se a métodos de usar proteínas de fusão ou componentes das mesmas.

Breve Descrição das Figuras

Figura 1 mostra um mapa de restrição de pMJ04.

Figura 2 mostra um mapa de restrição de pCaHj527.

Figura 3 mostra um mapa de restrição de pMT2188.

Figura 4 mostra um mapa de restrição de pCaHj568.

Figura 5 mostra um mapa de restrição de pMJ05.

Figura 6 mostra um mapa de restrição de pSMai130.

Figura 7 mostra a sequência de DNA e sequência de aminoácidos de uma sequência de sinal nativa de beta-glucosidase de *Aspergillus oryzae* (SEQ ID NOs 59 e 60).

Figura 8 mostra a sequência de DNA e sequência de aminoácidos de uma sequência de sinal de endoglucanase V de *Humicola insolens* (SEQ ID NOs 63 e 64).

Figura 9 mostra um mapa de restrição de pSMai135.

Figura 10 mostra um mapa de restrição de pSMai140.

Figura 11 mostra um mapa de restrição de pSaMe-F1.

Figura 12 mostra um mapa de restrição de pSaMe-FX.

Figura 13 mostra um mapa de restrição de pAILo47.

Figuras 14A, 14B, 14C, e 14D mostram uma sequência de DNA e sequência deduzida de aminoácidos da proteína de fusão variante de beta-glucosidase de *Aspergillus oryzae* (SEQ ID NOs 73 e 74, respectivamente).

Figuras 15A, 15B, 15C, e 15D mostram uma sequência de DNA e sequência deduzida de aminoácidos da proteína de fusão de beta-glucosidase de *Aspergillus oryzae* (SEQ ID NOs 75 e 76, respectivamente).

Definições

Atividade de endoglucanase: O termo "atividade de endoglucanase" é definido aqui como uma endo-1,4-beta-D-glucano 4-glucanohidrolase (E.C. No. 3.2.1.4) que catalisa a endohidrólise de ligações 1,4-beta-D-glicosídicas em celulose, derivados de celulose (como carboximetil celulose e hidroxietil celulose), liquenina, ligações beta-1,4 em misturas de beta-1,3 glucanos como beta-D-glucanos ou xiloglucanos de cereais, e outro material de planta contendo componentes celulósicos. Para os

fins da presente invenção, a atividade de endoglucanase é determinada usando hidrólise de carboximetil celulose (CMC) de acordo com o procedimento de Ghose, 1987, *Pure and Appl. Chem.* 59: 257 - 268. Uma unidade de atividade de endoglucanase é definida como 1,0 μmol de açúcares redutores produzidos por minuto a 50 °C, pH 4,8.

Atividade beta-glucosidase: O termo "beta-glucosidase" é definido aqui como uma beta-D-glucosídeo glucohidrolase (E.C. 32 1 21), que catalisa a hidrólise de resíduos de beta-D-glicose não redutores terminais com a liberação de beta-D-glicose. Para fins da presente invenção, atividade beta-glucosidase é determinada de acordo com o procedimento descrito por Venturi et al, 2002, *J Basic Microbiol* 42 55-66, exceto diferentes condições são empregadas como aqui descrito. Uma unidade de atividade beta-glucosidase é definida como 10 μmole de p-nitrofenol produzido por minuto a 50 °C pH 5 de 4 mM p-nitrofenil-beta-D-glucopiranosídeo como substrato em 100 mM citrato de sódio 0 01% TWEEN® 20.

Polipeptídeo de comprimento completo: O termo "comprimento completo polipeptídeo" é definido aqui como uma forma precursora de um polipeptídeo tendo atividade biológica, em que o precursor contém uma região de peptídeo sinal e alternativamente também uma região de propeptídeo, em que quando da secreção de uma célula, o peptídeo sinal é clivado e alternativamente também o propeptídeo é clivado dando um polipeptídeo com atividade biológica.

Peptídeo sinal: O termo "peptídeo sinal" é definido aqui como um peptídeo ligado em armação ao término amino de um polipeptídeo e dirige o polipeptídeo codificado em uma via secretória da célula.

Sequência de codificação de peptídeo sinal: O termo "sequência de codificação de peptídeo sinal" é definido aqui com uma região de codificação de peptídeo que codifica para uma sequência de aminoácidos ligada em armação ao término amino de um polipeptídeo codificado e dirige o

polipeptídeo codificado em uma via secretória da célula.

Propeptídeo: O termo "propeptídeo" é definido aqui como um peptídeo ligado em armação ao término amino de um polipeptídeo. O polipeptídeo resultante é conhecido como uma proenzima ou propolipeptídeo (ou um zimogênio em alguns casos) Um propolipeptídeo é geralmente inativo e pode ser convertido em um polipeptídeo ativo por clivagem catalítica ou autocatalítica do propeptídeo a partir do propolipeptídeo. Quando tanto as regiões do peptídeo sinal como do propeptídeo estão presentes no término amino de um polipeptídeo, a região de propeptídeo é ligada em armação ao término amino de um polipeptídeo e a região de peptídeo sinal é ligada em armação ao término amino da região de propeptídeo.

Sequência de codificação de propeptídeo: O termo "sequência de codificação de propeptídeo" é definido aqui como uma região de codificação de peptídeo que codifica para uma sequência de aminoácidos ligada em armação ao término amino de um polipeptídeo formando uma proenzima ou propolipeptídeo (ou um zimogênio em alguns casos).

Polipeptídeo maduro: O termo "polipeptídeo maduro" é definido aqui como um polipeptídeo tendo atividade de endoglucanase que está em sua forma final após tradução e quaisquer modificações pós-tradução tais, como processamento N-terminal, truncagem C-terminal, glicosilação, fosforilação, etc.

Domínio catalítico: O termo "domínio catalítico" é definido aqui como uma porção estrutural ou região da sequência de aminoácidos de uma endoglucanase ou um polipeptídeo tendo atividade biológica (por exemplo, atividade beta-glucosidase), que possui a atividade catalítica da endoglucanase ou do polipeptídeo tendo atividade biológica (por exemplo, beta- glucosidase). O domínio catalítico é também referido como a região de "núcleo" aqui.

Proteína de fusão: O termo "proteína de fusão" é definido

aqui como um polipeptídeo que demonstra atividade biológica e que compreende pelo menos tanto um domínio catalítico de endoglucanase como um domínio catalítico de um polipeptídeo tendo atividade biológica (por exemplo, beta- glucosidase).

5 **Proteína de fusão de beta-glucosidase:** O termo " proteína de fusão de beta-glucosidase " é definido aqui como um polipeptídeo que demonstra atividade beta-glucosidase e que compreende pelo menos tanto um domínio catalítico de beta-glucosidase como um domínio catalítico de endoglucanase.

10 **Componentes de uma proteína de fusão:** O termo "componentes de uma proteína de fusão" é definido aqui como fragmentos individuais (clivados) de uma proteína de fusão, em que cada fragmento tem atividade biológica e inclui ou pelo menos o domínio catalítico de uma endoglucanase e pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica ou pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica. Por exemplo, a presença de um sítio de clivagem, por exemplo, sítio Kex2, entre os componentes de pelo menos o domínio catalítico de uma endoglucanase e pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica de uma proteína de fusão pode resultar na produção de um polipeptídeo tendo atividade endoglucanase e outro polipeptídeo tendo atividade biológica.

20 **Componentes de uma proteína de fusão beta-glucosidase:**
O termo "componentes de uma proteína de fusão de beta-glucosidase " é definido aqui como fragmentos individuais (clivados) de uma proteína de fusão, em que cada fragmento tem atividade de beta-glucosidase e inclui ou pelo menos o domínio catalítico de uma endoglucanase e pelo menos o domínio catalítico da beta-glucosidase ou pelo menos o domínio catalítico da beta-glucosidase.

Por exemplo, a presença de um sítio de clivagem, por

exemplo, sítio Kex2, entre os componentes endoglucanase e beta-glucosidase da proteína de fusão pode resultar na produção de um polipeptídeo tendo atividade endoglucanase e outro polipeptídeo tendo atividade beta-glucosidase.

5 **Módulo de ligação de carboidrato.** O termo "módulo de ligação de carboidrato" (CBM)" é definido aqui como uma porção da sequência de aminoácidos de uma endoglucanase (celulase) que está envolvida na ligação de uma endoglucanase a celulose (lignocelulose). Os módulos de ligação de carboidratos geralmente funcionam por ligação não
10 covalente da endoglucanase em celulose, um derivado de celulose, ou um equivalente de polissacarídeo da mesma. CBMs tipicamente funcionam independentes de um domínio catalítico.

Construção de proteína de fusão. O termo "construção de proteína de fusão" refere-se a uma construção de ácido nucleico que é
15 composta de diferentes genes ou porções dos mesmos em ligação operativa. Os componentes incluem a partir da extremidade 5' uma molécula de DNA codificando pelo menos um domínio catalítico de endoglucanase e uma molécula de DNA codificando pelo menos um domínio catalítico de um polipeptídeo tendo atividade biológica.

20 **Construção de fusão de Beta-glucosidase:** O termo " construção de fusão de beta-glucosidase " refere-se a uma construção de ácido nucleico que é composta de diferentes genes ou porções dos mesmos em ligação operativa. Os componentes incluem a partir de uma extremidade 5' uma molécula de DNA codificando pelo menos um domínio catalítico de
25 endoglucanase e uma molécula de DNA codificando pelo menos um domínio catalítico de beta-glucosidase.

Polipeptídeo isolado: O termo "polipeptídeo isolado" como usado aqui refere-se a um polipeptídeo que é pelo menos 20% puro, preferivelmente pelo menos 40% puro, mais preferivelmente pelo menos 60%

puro, ainda mais preferivelmente pelo menos 80% puro, o mais preferivelmente pelo menos 90% puro, e ainda o mais preferivelmente pelo menos 95% puro, como determinado por SDS-PAGE.

Polipeptídeo substancialmente puro: O termo "polipeptídeo substancialmente puro " denota aqui uma preparação de polipeptídeo que contém no máximo 10%, preferivelmente no máximo 8%, mais preferivelmente no máximo 6%, mais preferivelmente no máximo 5%, mais preferivelmente no máximo 4%, mais preferivelmente no máximo 3%, ainda mais preferivelmente no máximo 2%, o mais preferivelmente no máximo 1%, e ainda o mais preferivelmente no máximo 0,5% em peso de outro material de polipeptídeo que é associado de modo nativo ou recombinante. Prefere-se, assim, que o polipeptídeo substancialmente puro seja pelo menos 92% puro, preferivelmente pelo menos 94% puro, mais preferivelmente pelo menos 95% puro, mais preferivelmente pelo menos 96% puro, mais preferivelmente pelo menos 96% puro, mais preferivelmente pelo menos 97% puro, mais preferivelmente pelo menos 98% puro, ainda mais preferivelmente pelo menos 99%, o mais preferivelmente pelo menos 99,5% puro, e ainda o mais preferivelmente 100% puro em peso do material de polipeptídeo total presente na preparação. Os polipeptídeos da presente invenção estão preferivelmente em uma forma substancialmente pura, isto é, a preparação de polipeptídeo é essencialmente livre de outro material de polipeptídeo como o qual é associado de modo nativo ou recombinante. Isto pode ser obtido, por exemplo, por preparação do polipeptídeo por meio de métodos recombinantes bem conhecidos ou por métodos de purificação clássicos.

Identidade: O parentesco entre duas sequências de aminoácido ou entre duas sequências de nucleotídeo é descrito pelo parâmetro "identidade".

Para os fins da presente invenção, o grau de identidade entre duas sequências de aminoácido é determinado usando o algoritmo

Needleman-Wunsch (Needleman e Wunsch, 1970, *J. Mol. Biol.* 48: 443 - 453) como implementado no programa Needle de EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al, 2000, *Trends in Genetics* 16: 276-277), preferivelmente versão 3.0.0 ou posterior. Os parâmetros opcionais usados são penalidade de abertura de espaço de 10, penalidade de extensão de espaço de 0,5, e a matriz de substituição EBLOSUM62 (EMBOSS versão de BLOSUM 62). A saída de Needle rotulada "identidade mais longa" (obtida usando a opção -nobrief) é usada como a porcentagem de identidade e é calculada como a seguir:

$$\text{10} \quad \frac{(\text{Resíduos idênticos} \times 100)}{(\text{Comprimento de Alinhamento} - \text{Número de espaços em alinhamento})}$$

Para os fins da presente invenção, o grau de identidade entre duas sequências de nucleotídeo é determinado usando o algoritmo Needleman-Wunsch (Needleman e Wunsch, 1970, supra) como implementado no programa de Needle de EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al, 2000) preferivelmente versão 3.0.0 ou posterior. Os parâmetros opcionais usados são penalidade de abertura de espaço de 10, penalidade de extensão de espaço de 0,5, e a matriz de substituição de EDNAFULL (Versão EMBOSS, de NCBI NUC4.4). A saída de Needle rotulada "identidade mais longa" (obtida usando a opção nobrief) é usada como a porcentagem de identidade e é calculada como a seguir:

$$\text{20} \quad \frac{(\text{deoxirribonucleotídeos idênticos} \times 100)}{(\text{Comprimento de Alinhamento} - \text{Número total de Espaços em Alinhamento})}$$

25 **Fragmento de polipeptídeo:** O termo "fragmento de polipeptídeo" é definido aqui como um polipeptídeo tendo um ou mais aminoácidos deletados do término amino e/ou carboxila de uma proteína de fusão (por exemplo, proteína de fusão de beta-glucosidase) ou componentes dos mesmos, em que o fragmento tem atividade biológica (por exemplo,

atividade beta-glicosidase)

Subsequência: O termo "subsequência" é definido aqui como uma sequência de nucleotídeos tendo um ou mais nucleotídeos deletado da extremidade 5' e/ou 3' de um polinucleotídeo, em que a subsequência codifica um fragmento de polipeptídeo tendo atividade biológica, por exemplo, atividade beta-glicosidase ou atividade endoglucanase

Polinucleotídeo isolado: O termo "polinucleotídeo isolado" como usado aqui refere-se a um polinucleotídeo que é pelo menos 20% puro, preferivelmente pelo menos 40% puro, mais preferivelmente pelo menos 60% puro, ainda mais preferivelmente pelo menos 80% puro, o mais preferivelmente pelo menos 90% puro, e ainda o mais preferivelmente pelo menos 95% puro, como determinado por eletroforese de agarose.

Polinucleotídeo substancialmente puro: O termo "polinucleotídeo substancialmente puro" como usado aqui refere-se a uma preparação de polinucleotídeo livre de outros nucleotídeos estranhos ou indesejados e em uma forma apropriada para usar em sistemas de produção de proteína geneticamente engenheirados. Assim, um polinucleotídeo substancialmente puro contém no máximo 10%, preferivelmente no máximo 8%, mais preferivelmente no máximo 6%, mais preferivelmente no máximo 5%, mais preferivelmente no máximo 4%, mais preferivelmente no máximo 3%, ainda mais preferivelmente no máximo 2%, o mais preferivelmente no máximo 1%, e ainda o mais preferivelmente no máximo 0,5% em peso de outro material de polinucleotídeo com o qual ele é naturalmente ou recombinantemente associado. Um polinucleotídeo substancialmente puro pode, no entanto, incluir regiões 5' e 3' de ocorrência natural não traduzidas, como promotores e terminadores. Prefere-se que o polinucleotídeo substancialmente puro seja pelo menos 90% puro, preferivelmente pelo menos 92% puro, mais preferivelmente pelo menos 94% puro, mais preferivelmente pelo menos 95% puro, mais preferivelmente pelo menos 96%

puro, mais preferivelmente pelo menos 97% puro, ainda mais preferivelmente pelo menos 98% puro, o mais preferivelmente pelo menos 99%, e ainda o mais preferivelmente pelo menos 99,5% puro em peso. Os polinucleotídeos estão preferivelmente em uma forma substancialmente pura, isto é, que a
5 preparação de polinucleotídeo seja essencialmente livre de outro material de polinucleotídeo com o qual ele é naturalmente ou recombinantemente associado. Os polinucleotídeos podem ser de origem genômica, cDNA, RNA, semi-sintética, sintética, ou quais combinações dos mesmos.

cDNA: O termo "cDNA" é definido aqui como uma molécula
10 de DNA que pode ser preparada por transcrição reversa de uma molécula de mRNA maduro, emendada, obtida de uma célula eucariótica. cDNA não tem as sequências de íntrons que estão geralmente presentes no DNA genômico correspondente. O transcrito de RNA primário inicial é um precursor para mRNA que é processado através de uma série de etapas antes de aparecer
15 como um mRNA emendado maduro. Estas etapas incluem a remoção de sequências de intron por um processo chamado "emenda". cDNA derivado de mRNA não apresenta, assim, quaisquer sequências de intron.

Construção de ácido nucléico: O termo "construção de ácido nucléico" como usado aqui refere-se a uma molécula de ácido nucléico, de
20 filamento único ou duplo, que é isolada de um gene de ocorrência natural ou que é modificada para conter segmentos de ácidos nucleicos em um modo que não iria de outra forma existir na natureza. O termo construção de ácido nucléico é sinônimo do termo "cassete de expressão" quando a construção de ácido nucléico contém as sequências de controle requeridas para expressão de
25 uma sequência de codificação.

Sequência de controle: O termo "sequência de controle" é definido aqui para incluir todos componentes, que são necessários ou vantajosos para a expressão de um polinucleotídeo codificando um polipeptídeo. Cada sequência de controle pode ser nativa ou estranha para a

sequência de nucleotídeos codificando o polipeptídeo ou nativa ou estranha para cada outro. Tais sequências de controle incluem, mas não são limitados a um líder, sequência de poliadenilação, sequência de propetídeo, promotor, sequência de peptídeo sinal, e terminador de transcrição. Em um mínimo, as

5 sequências de controle incluem um promotor, e sinais de parada transcripcional e translacional. As sequências de controle podem ser providas com ligadores para o fim de introduzir sítios de restrição específicos facilitando a ligação das sequências de controle com a região codificada da sequência de nucleotídeos codificando um polipeptídeo.

10 **Ligado operativamente:** O termo "ligado operativamente" denota aqui uma configuração em que uma sequência de controle é colocada em uma posição apropriada relativa na sequência de codificação da sequência de polinucleotídeo de modo que a sequência de controle dirige a expressão da

15 sequência de codificação de um polipeptídeo. Além disso, o termo "ligado operativamente " também se refere a dois polinucleotídeos que são ligados ou fusionados, que são expressados juntos como uma proteína fusionada ou de fusão.

Sequência de codificação: Quando usado aqui o termo "sequência de codificação" significa uma sequência de nucleotídeo, que

20 diretamente especifica a sequência de aminoácido de seu produto de proteína. Os limites da sequência de codificação geralmente são determinados por uma matriz de leitura aberta, que geralmente começa com o códon de partida ATG ou códons de partida alternativos como GTG e TTG e terminado com um códon de parada como TAA, TAG, e TGA. A sequência de codificação pode

25 ser uma sequência de DNA, cDNA, ou de nucleotídeo recombinante.

Expressão: O termo "expressão" inclui qualquer etapa envolvida na produção do polipeptídeo incluindo, mas não limitado a transcrição, modificação pós-transcripcional, tradução, modificação pós-translacional, e secreção.

Vetor de expressão: O termo "vetor de expressão" é definido aqui como uma molécula de DNA linear ou circular que compreende um polinucleotídeo codificando um polipeptídeo da invenção, e que é ligado operativamente a nucleotídeos adicionais que provê sua expressão.

5 **Célula hospedeira:** O termo "célula hospedeira", como usado aqui, inclui qualquer tipo de célula que é suscetível para transformação, transfecção, transdução, e outros com uma construção de ácido nucléico ou vetor de expressão compreendendo um polinucleotídeo.

10 **Variante:** Quando usado aqui, o termo "variante" significa um polipeptídeo tendo atividade biológica produzida por um organismo expressando uma sequência de nucleotídeo modificada, por exemplo SEQ ID NO: 25, ou uma sequência homóloga da mesma, ou a região de codificação madura da mesma. A sequência de nucleotídeo modificada é obtida através de intervenção humana por modificação de uma sequência de nucleotídeos, por
15 exemplo SEQ ID NO: 23, ou uma sequência homóloga da mesma, ou a região de codificação madura da mesma. A modificação pode ser uma substituição, uma deleção e/ou uma inserção de um ou mais aminoácidos assim como uma substituição de uma ou mais cadeias laterais de aminoácidos.

Descrição Detalhada da Invenção

20 A presente invenção refere-se a métodos para produzir um polipeptídeo secretado tendo atividade biológica, compreendendo (a) transformar uma célula hospedeira fúngica com uma construção de proteína de fusão codificando a proteína de fusão, em que a construção de proteína de fusão compreende (i) um primeiro polinucleotídeo compreendendo uma
25 sequência de nucleotídeos codificando um peptídeo sinal, (ii) um segundo polinucleotídeo compreendendo uma sequência de nucleotídeos codificando pelo menos um domínio catalítico de uma endoglucanase ou uma porção da mesma, e (iii) um terceiro polinucleotídeo compreendendo uma sequência de nucleotídeos codificando pelo menos um domínio catalítico de um

polipeptídeo tendo atividade biológica ou uma porção do mesmo, em que o peptídeo sinal e pelo menos o domínio catalítico da endoglucanase ou a porção do mesmo aumenta a secreção do polipeptídeo tendo atividade biológica ou a porção do mesmo comparado com a ausência de pelo menos o domínio catalítico da endoglucanase ou a porção do mesmo, (b) cultivar a célula hospedeira fúngica transformada sob condições apropriadas para a produção da proteína de fusão, e (c) recuperar a proteína de fusão, um componente da mesma, ou uma combinação da mesma, a partir do meio de cultivo, em que a proteína de fusão ou o componente da mesma tem atividade biológica.

Em um aspecto preferido, a extremidade 3' do primeiro polinucleotídeo é ligada operativamente à extremidade 5' do segundo polinucleotídeo e a extremidade 3' do segundo polinucleotídeo é ligada operativamente à extremidade 5' do terceiro polinucleotídeo. Em outro aspecto preferido, a extremidade 3' do primeiro polinucleotídeo é ligada operativamente à extremidade 5' do terceiro polinucleotídeo e a extremidade 3' do terceiro polinucleotídeo é ligada operativamente à extremidade 5' do segundo polinucleotídeo para codificar a proteína de fusão.

A proteína de fusão é produzida por fusão de uma sequência de nucleotídeos codificando um polipeptídeo tendo atividade biológica ou uma porção do mesmo para uma sequência de nucleotídeos codificando um polipeptídeo tendo atividade endoglucanase ou uma porção da mesma e uma sequência de nucleotídeos codificando um peptídeo sinal ligada operativamente à sequência de nucleotídeos codificando o polipeptídeo tendo atividade endoglucanase ou uma porção da mesma. Técnicas para produzir proteínas de fusão são conhecidas na técnica, e incluem, por exemplo, ligação das sequências de codificação codificando os polipeptídeos de modo que estejam em armação e expressão do polipeptídeo fusionado é ainda controlar o(s) mesmo(s) promotor (es) e terminador. As proteínas de fusão também

podem ser construídas usando tecnologia de inteína em que as fusões são criadas pós-translacionalmente (Cooper et al, 1993, *EMBO J* 12 2575-2583, Dawson et al, 1994, *Science* 266 776-779).

5 A proteína de fusão tendo atividade biológica compreendendo um peptídeo sinal, pelo menos o domínio catalítico de uma endoglucanase ou uma porção da mesma, e pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica ou uma porção do mesmo, aumenta secreção de uma proteína de fusão comparado com a ausência de pelo menos o domínio catalítico da endoglucanase ou uma porção da mesma. O aumento na secreção
10 de uma proteína de fusão tendo atividade biológica é pelo menos 5%, preferivelmente pelo menos 10%, mais preferivelmente pelo menos 25%, ainda mais preferivelmente pelo menos 50%, mais preferivelmente pelo menos 100%, ainda mais preferivelmente pelo menos 150%, ainda mais preferivelmente pelo menos 200%, o mais preferivelmente pelo menos 500%,
15 e ainda o mais preferivelmente pelo menos 1000% comparado com a ausência de pelo menos o domínio catalítico da endoglucanas.

Em cada um dos aspectos preferidos abaixo, os componentes de uma construção de proteína de fusão (construção de ácido nucleico) são ligados operativamente a partir da extremidade 5' à extremidade 3' da
20 construção.

Em um aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio
25 catalítico de uma endoglucanase, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de

nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo

compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

5 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo compreendendo uma
10 sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo
15 compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

 Em outro aspecto preferido, a construção de proteína de fusão
20 compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento
25 completo de um polipeptídeo tendo atividade biológica.

 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo

de comprimento completo de um polipeptídeo tendo atividade biológica, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

5 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo compreendendo uma
10 sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo
15 de comprimento completo de um polipeptídeo tendo atividade biológica, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de
20 nucleotídeos codificando um polipeptídeo de comprimento completo of uma endoglucanase (peptídeo sinal e polipeptídeo maduro), e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão
25 compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo of uma endoglucanase (peptídeo sinal e polipeptídeo maduro).

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

10 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

20 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

25 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo

compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos

codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

5 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

10 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

20 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase. Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de

25

nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e
5 um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de
10 nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, e um polinucleotídeo compreendendo
15 uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de
20 nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

25 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, um polinucleotídeo

compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

5 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um
10 polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo
15 compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

20 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, um polinucleotídeo compreendendo uma
25 sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando

um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica; um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica. Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, um polinucleotídeo compreendendo uma sequência de

nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

5 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de
10 carboidrato, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

15 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato,
20 um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica.

25 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de

ligação de carboidrato, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

5 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase, e um polinucleotídeo codificando um
10 ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

15 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polinucleotídeo
20 codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um domínio catalítico de uma endoglucanase.

25 Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase, e um polinucleotídeo codificando um ligador

e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a construção de proteína de fusão compreende um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polinucleotídeo codificando um ligador e/ou um polinucleotídeo codificando um módulo de ligação de carboidrato, um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando outro peptídeo sinal, e um polinucleotídeo compreendendo uma sequência de nucleotídeos codificando um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, para cada um dos aspectos preferidos acima, os polinucleotídeos podem codificar uma porção do domínio catalítico, o polipeptídeo maduro, ou o polipeptídeo de comprimento completo de uma endoglucanase ou um polipeptídeo tendo atividade biológica. A porção da endoglucanase pode ter ou não pode ter atividade endoglucanase. Em um aspecto mais preferido, a porção da endoglucanase tem atividade endoglucanase.

Em cada um dos aspectos preferidos acima, os componentes das construções de proteína de fusão ainda compreendem uma região de promotor e/ou uma região de terminador.

Endoglucanases e polinucleotídeos das mesmas

Um polinucleotídeo codificando um domínio catalítico, polipeptídeo maduro, ou polipeptídeo de comprimento completo de uma endoglucanase, ou porções da mesma, pode ser obtido a partir de qualquer

organismo. Para fins da presente invenção, o termo "polipeptídeo" será entendido para incluir um polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico, ou porções ou fragmentos dos mesmos que tem atividade. O termo "obtido a partir de", como usado aqui em
5 conexão com uma dada fonte deve significar que o polipeptídeo codificado por uma sequência de nucleotídeos é produzido pela fonte ou pela cepa em que a sequência de nucleotídeos a partir da fonte foi inserida.

Muitas endoglucanases têm uma estrutura de múltiplos domínios consistindo de um domínio catalítico separado de um domínio de
10 ligação de carboidratos (CBM) peptídeo ligador (Suurnakki et al, 2000, *Celulose* 7 189-209) O domínio catalítico contém o sítio ativo enquanto o CBM interage com celulose por ligação da enzima ao mesmo (van Tilbeurgh et al, 1986, *FEBS Letters* 204 223-227, Tomme et al, 1988, *European Journal of Biochemistry* 170 575-581)

15 Um polinucleotídeo codificando um polipeptídeo tendo atividade endoglucanase pode ser obtido a partir de um gene codificando um polipeptídeo bacteriano. Por exemplo, o polipeptídeo pode ser um polipeptídeo bacteriano Gram positivo, incluindo, mas não limitado a, polipeptídeo de *Bacillus*, *Streptococcus*, *Streptomyces*, *Staphylococcus*,
20 *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus*, ou *Oceanobacillus* polipeptídeo, por exemplo, um polipeptídeo de *Bacillus alkalophilus*, *Bacillus amiloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*. *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*. *Bacillus*
25 *thuringiensis*, *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus ubens*, e *Streptococcus equi subsp Zooepidemicus*, *Streptomyces lividans*, ou *Streptomyces murinus*, ou um polipeptídeo bacteriano Gram negativo, incluindo, mas não limitado a, um polipeptídeo de *E coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*,

Flavobacterium, Fusobacterium, Ilyobacter, Neisseria, ou Ureaplasma.

Os exemplos de endoglucanases bacterianas que podem ser usadas como fontes para os polinucleotídeos nos métodos da presente invenção, incluem, mas não são limitados a endoglucanase de *Acidothermus*
 5 *celulolyticus* (WO 91/05039, WO 93/15186, Patente US 5.275.944, WO 96/02551, Patente US 5.536.655, WO 00/70031, WO 05/093050), endoglucanase III de *Thermobifida fusca* (WO 05/093050), e endoglucanase V de *Thermobifida fusca* (WO 05/093050)

Um polinucleotídeo codificando um polipeptídeo tendo
 10 atividade endoglucanase pode ser obtido a partir de um gene codificando um polipeptídeo fúngico, e mais preferivelmente polipeptídeo de levedura, como um polipeptídeo de *Candida, Kluyveromyces, Pichia, Saccharomyces, Schizosaccharomyces*, ou *Yarrowia*, ou mais preferivelmente um polipeptídeo de fungo filamentoso como um polipeptídeo de *Acremonium, Aspergillus,*
 15 *Aureobasidium, Cladorrhinum, Cryptococcus, Filibasidium, Fusarium, Humicola, Magnaporthe, Mucor, Mycelophthora, Neocallimastix, Neurospora, Paealomyces, Pemallium, Pimmyces, Schizophyllum, Talaromyces, Thermoascus, Thielavia, Tolypocladium,* ou *Trichodema.*

Em um aspecto preferido, um polinucleotídeo codificando um
 20 polipeptídeo tendo atividade endoglucanase pode ser obtido a partir de um gene codificando um polipeptídeo de *Saccharomyces carlsbergensis, Saccharomyces cerevisiae, Saccharomyces diastaticus, Saccharomyces douglasii, Saccharomyces kluyveri, Saccharomyces norbensis,* ou *Saccharomyces oviformis*

Em outro aspecto preferido, um polinucleotídeo codificando
 25 um polipeptídeo tendo atividade endoglucanase pode ser obtido a partir de um gene codificando um polipeptídeo de *Aspergillus aculeatus, Aspergillus awamori, Aspergillus fumigatus, Aspergillus foetidus, Aspergillus japonicus, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Cladorrhinum*

foecundissimum, *Fusarium bactrodioides*, *Fusarium cerealis*, *Fusarium*
crookwellense, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium*
graminum, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium*
oxysporum, *Fusarium reticulatum*, *Fusarium riseum*, *Fusarium sambuanum*,
5 *Fusarium sarcochmum*, *Fusarium spomtrchioides*, *Fusarium sulphureum*,
Fusarium torulosum, *Fusarium tnchothecioides*, *Fusarium venenatum*,
Humicola insolens, *Humicola lanuginosa*, *Mucor miehei*, *Mycehophthora*
thermophila, *Neurospora crassa*, *Pemallium purpungenum*, *Thielavia*
achromatica, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia*
10 *australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*,
Thielavia peruviana, *Thielavia spededonium*, *Thielavia setosa*, *Thielavia sub*
thermophila, *Thielavia terrestris*, *Thielavia temcola*, *Thielavia thermophila*,
Thielavia vanospora, *Thielavia warengii*, *Trichoderma harzianum*,
Trichoderma komngii, *Trichoderma longibrachiatum*, *Trichoderma reesei*, ou
15 *Trichoderma vinde*.

Exemplos de endoglucanases fúngicas que podem ser usadas
 como fontes para os polinucleotídeos nos métodos da presente invenção,
 incluem, mas não são limitados, a *Trichoderma reesei* EG1 (Penttila et al,
 1986, *Gene* 45 253-263, GENBANK™ número de acesso M15665),
 20 *Trichoderma reesei* EG2 (Saloheimo, et al, 1988, *Gene* 63 11- 22,
 GENBANK™ número de acesso M19373), *Trichoderma reesei* EG3 (Okada
 et al, 1988, *Appl Environ Microbiol* 64 555-563, GENBANK™ número de
 acesso AB003694), *Trichoderma reesei* EG4 (Saloheimo et al, 1997, *Eur. J.*
Biochem 249 584-591, GENBANK™ número de acesso Y11113), e
 25 *Trichoderma reesei* EG5 (Saloheimo et al, 1994, *Molecular Microbiology* 13
 219-228, GENBANK™ número de acesso Z33381), endoglucanase de
Aspergillus aculeatus (Ooi et al, 1990, *Nucleic Acid Research* 18 5884),
Aspergillus kawachn (Sakamoto et al, 1995, *Current Genetics* 27 435-439),
 endoglucanase de *Erwinia carotovara* (Saarilahti et al, 1990, *Gene* 90 9-14),

Fusarium oxysporum (GENBANK™ número de acesso L29381), *Humicola grisea var thermoidea* (GENBANK™ número de acesso AB003107), *Melanocarpus albomyces* (GENBANK™ número de acesso MAL515703), e *Neurospora crassa* (GENBANK™ número de acesso XM_324477).

5 Outras endoglucanases são descritas em numerosas famílias de Glicosil Hidrolase usando a classificação de acordo com Henrissat B., 1991, A classification of glicosil hidrolases based on amino-acid sequence similarities, *Biochem J.* 280 309-316, e Henrissat B., e Bairoch A., 1996, Updating the sequence-based classification of glicosil hidrolases, *Biochem J.*
10 316 695-696.

 As técnicas usadas para isolar ou clonar um polinucleotídeo codificando um polipeptídeo tendo atividade endoglucanase são conhecidas na técnica e incluem isolamento do DNA genômico, preparação do cDNA ou uma combinação dos mesmos. A clonagem dos polinucleotídeos a partir deste
15 DNA genômico pode ser efetuado, por exemplo, usando a reação de cadeia polimerase bem conhecida (PCR) ou triagem de anticorpos de bibliotecas de expressão para detectar os fragmentos de DNA clonados com aspectos estruturais compartilhados. Ver, por exemplo, Innis et al, 1990, *PCR A Guide to Methods and Application*, Academic Press, New York. Outros
20 procedimentos de amplificação de ácido nucleico como reação de cadeia ligase (LCR), transcrição ativada ligada (LAT) e amplificação com base em sequência de nucleotídeos (NASBA) podem ser usados.

 Será entendido que para as espécies acima mencionadas, a invenção engloba tanto os estados perfeitos como imperfeitos, e outros
25 equivalentes taxonômicos, por exemplo anamorfos, sem levar em conta o nome da espécie pelo qual são conhecidos. Os versados na arte irão prontamente reconhecer a identidade de equivalentes apropriados.

 As cepas destas espécies são prontamente acessíveis ao público em várias coleções de culturas, como o American Type Culture

Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

5 Em um aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase I.

10 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase II.

15 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase III.

20 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase IV.

25 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V. Em um aspecto mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens*. Em um aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou

uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* que codifica o polipeptídeo de SEQ ID NO 2. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* que codifica o polipeptídeo de comprimento completo de SEQ ID NO 2. Em um outro aspecto o mais preferido, o polipeptídeo maduro da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* que codifica o polipeptídeo maduro de SEQ ID NO 2. Em um outro aspecto o mais preferido, o domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* que codifica o domínio catalítico de SEQ ID NO 2. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* compreendendo SEQ ID NO 1. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* compreendendo SEQ ID NO 1. Em um outro aspecto o mais preferido, o polipeptídeo maduro da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* compreendendo o polipeptídeo maduro codificando sequência de SEQ ID NO 1. Em um outro aspecto o mais preferido, o domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase V de *Humicola insolens* compreendendo o domínio catalítico

codificando sequência de SEQ ID NO 1.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase VI.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase da Família 5. Em um aspecto mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Mycehophthora thermophila* CBS 117 65. Em um aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Mycehophthora thermophila* CBS 117 65 que codifica o polipeptídeo de SEQ ID NO 4. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Myceliophthora thermophila* CBS 117 65 compreendendo SEQ ID NO 3. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de basidiomicete CBS 495 95. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de

basidiomicete CBS 495 95 que codifica o polipeptídeo de SEQ ID NO 6. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de basidiomicete CBS 495 95 compreendendo SEQ ID NO 5.

5 Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de basidiomicete CBS 494 95. Em outro aspecto o

10 mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de basidiomicete CBS 494 95 que codifica o polipeptídeo de SEQ ID NO 8. Em outro aspecto o mais preferido, polipeptídeo de

15 comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de basidiomicete CBS 494 95 compreendendo SEQ ID NO 7.

Em outro aspecto preferido, o polipeptídeo de comprimento

20 completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase Família 6. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo é codificado por um

25 polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL6B. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia*

terrestris NRRL 8126 CEL6B que codifica o polipeptídeo de SEQ ID NO 10. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL6B compreendendo

5 SEQ ID NO 9. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia*

10 *terrestris* NRRL 8126 CEL6C Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL6C que codifica o polipeptídeo de SEQ ID NO 12.

15 Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL6C compreendendo SEQ ID NO 11.

20 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de Família 7. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou

25 domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7C. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um

polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7C que codifica o polipeptídeo de SEQ ID NO 14. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou

5 uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7C compreendendo SEQ ID NO 13. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um

10 polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7E. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia*

15 *terrestris* NRRL 8126 CEL7E que codifica o polipeptídeo de SEQ ID NO 16. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7E compreendendo

20 SEQ ID NO 15. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7F. Em outro aspecto o mais preferido, o

25 polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7F que codifica o polipeptídeo de SEQ ID NO 18. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo,

polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7F compreendendo SEQ ID NO 17. Em um outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase *Cladorrhinum foecundissimum* ATCC 62373 CEL7A. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Cladorrhinum foecundissimum* ATCC 62373 CEL7A que codifica o polipeptídeo de SEQ ID NO 20. Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Cladorrhinum foecundissimum* ATCC 62373 CEL7A compreendendo SEQ ID NO 19.

Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Trichoderma reesei* cepa No VTT-D-80133 que codifica o polipeptídeo de SEQ ID NO 22 (GENBANK™ número de acesso M15665).

Em outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase de *Trichoderma reesei* cepa No VTT-D- 80133 compreendendo SEQ ID NO 21

(GENBANK™ número de acesso M 15665).

5 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase da Família 9.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase da Família 12.

10 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase da Família 45.

15 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene de endoglucanase da Família 74.

20 Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo obtido a partir de um gene codificando um polipeptídeo homólogo compreendendo uma sequência de aminoácidos que tem um grau de identidade para as sequências de aminoácidos do polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico de SEQ ID NO 2, SEQ ID NO 4, SEQ ID NO 6, ou SEQ ID NO 8, SEQ ID NO 10, SEQ ID NO 12, SEQ ID NO 14, SEQ ID NO 16, SEQ ID NO 18, SEQ ID NO 20, ou SEQ ID NO 22 de pelo menos 60%, preferivelmente pelo menos 65%, mais preferivelmente pelo menos 70%, mais preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, mais preferivelmente pelo menos 85%, ainda mais

preferivelmente pelo menos 90%, o mais preferivelmente pelo menos 95%, e ainda o mais preferivelmente 96%, 97%, 98% ou 99%, que tem atividade de endoglucanase. Em um aspecto preferido, o polipeptídeo homólogo tem uma sequência de aminoácidos que difere por dez aminoácidos, preferivelmente
5 por cinco aminoácidos, mais preferivelmente por quatro aminoácidos, ainda mais preferivelmente por três aminoácidos, o mais preferivelmente por dois aminoácidos, e ainda o mais preferivelmente por um aminoácido do polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico de SEQ ID NO 2, SEQ ID NO 4, SEQ ID NO 6, ou SEQ ID NO 8,
10 SEQ ID NO 10, SEQ ID NO 12, SEQ ID NO 14, SEQ ID NO 16, SEQ ID NO 18, SEQ ID NO 20, ou SEQ ID NO 22.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da endoglucanase, ou uma porção do mesmo, é codificado por um polinucleotídeo compreendendo
15 uma sequência de nucleotídeos que hibridiza sob condições de estringência muito baixas, preferivelmente condições de estringência baixa, mais preferivelmente condições de estringência média, mais preferivelmente condições de estringência média elevada, ainda mais preferivelmente condições de estringência alta, e o mais preferivelmente condições de
20 estringência muito alta com (i) SEQ ID NO 1, SEQ ID NO 3, SEQ ID NO 5, ou SEQ ID NO 7, SEQ ID NO 9, SEQ ID NO 11, SEQ ID NO 13, SEQ ID NO 15, SEQ ID NO 17, SEQ ID NO 19 ou SEQ ID NO 21, (ii) a sequência de cDNA contida em SEQ ID NO 1, SEQ ID NO 3, SEQ ID NO 5, ou SEQ ID NO 7, SEQ ID NO 9, SEQ ID NO 11, SEQ ID NO 13, SEQ ID NO 15,
25 SEQ ID NO 17, SEQ ID NO 19, ou SEQ ID NO 21, (iii) uma subsequência de (i) ou (ii) ou (iv) de um filamento complementar de (i), (ii) ou (iii), (J Sambrook, E. F. Fritsch, e T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2d ed. Cold Spring Harbor, New York) A subsequência de SEQ ID NO 1, SEQ ID NO 3, SEQ ID NO 5, ou SEQ ID NO 7, SEQ ID

NO 9, SEQ ID NO 11, SEQ ID NO 13, SEQ ID NO 15, SEQ ID NO 17, SEQ ID NO 19, ou SEQ ID NO 21 contém pelo menos 100 nucleotídeos contíguos ou preferivelmente pelo menos 200 nucleotídeos contíguos.

A sequência de nucleotídeos de SEQ ID NO 1, SEQ ID NO 3, SEQ ID NO 5, ou SEQ ID NO 7, SEQ ID NO 9, SEQ ID NO 11, SEQ ID NO 13, SEQ ID NO 15, SEQ ID NO 17, SEQ ID NO 19, ou SEQ ID NO 21, ou a subsequência das mesmas, assim como a sequência de aminoácidos de SEQ ID NO 2, SEQ ID NO 4, SEQ ID NO 6, ou SEQ ID NO 8, SEQ ID NO 10, SEQ ID NO 12, SEQ ID NO 14, SEQ ID NO 16, SEQ ID NO 18, SEQ ID NO 20, ou SEQ ID NO 22, ou um fragmento das mesmas, pode ser usada para projetar uma sonda de ácido nucleico para identificar e clonar polipeptídeos codificando DNA tendo atividade de endoglucanase a partir de cepas de diferentes gêneros ou espécies de acordo com métodos bem conhecidos na arte. Em particular, estas sondas podem ser usadas para hibridização com o cDNA ou genômico do gênero ou espécie de interesse, seguindo procedimentos de Southern blotting padrões, a fim de identificar e isolar o gene correspondente ai. Estas sondas podem ser consideravelmente mais curtas do que a sequência completa, mas devem ser pelo menos 14, preferivelmente pelo menos 25, mais preferivelmente pelo menos 35, e o mais preferivelmente pelo menos 70 nucleotídeos em comprimento. No entanto, prefere-se que a sonda de ácido nucleico tenha pelo menos 100 nucleotídeos em comprimento. Por exemplo, a sonda de ácido nucleico pode ter pelo menos 200 nucleotídeos, preferivelmente pelo menos 300 nucleotídeos, mais preferivelmente pelo menos 400 nucleotídeos, ou o mais preferivelmente pelo menos 500 nucleotídeos de comprimento. Sondas ainda mais longas podem ser usadas, por exemplo, sondas de ácido nucleico que tem pelo menos 600 nucleotídeos, pelo menos preferivelmente pelo menos 700 nucleotídeos, mais preferivelmente pelo menos 800 nucleotídeos, ou o mais preferivelmente pelo menos 900 nucleotídeos em comprimento. Tanto sondas de DNA como de

RNA podem ser usadas. As sondas são tipicamente rotuladas para detecção do gene correspondente (por exemplo com ^{32}P , ^3H , ^{35}S , biotina, ou avidina). Estas sondas estão englobadas na presente invenção.

5 Uma biblioteca de DNA genômico ou cDNA preparada a partir destes outros organismos pode, assim, ser triada para DNA que hibridiza com as sondas descritas acima e codifica um polipeptídeo tendo atividade de endoglucanase. DNA genômico ou outro a partir destes outros organismos pode ser separado por eletroforese de gel agarose ou poliacrilamida ou outras técnicas de separação. O DNA de bibliotecas ou o
10 DNA separado pode ser transferido e imobilizado em nitrocelulose ou outro material de veículo apropriado. A fim de identificar um clone ou DNA que é homólogo com as seqüências acima, o material veículo é usado em Southern Blot.

15 Para fins da presente invenção, a hibridização indica que a seqüência nucleotídica hibridiza em uma sonda de ácido nucleico rotulada correspondendo a uma das seqüências de nucleotídeos descritas acima sob condições de estringência muito baixa a muito alta. As moléculas em que as sondas de ácido nucleico hibridizam sob estas condições podem ser detectadas usando filme de raio-X.

20 Para sondas longas de pelo menos 100 nucleotídeos de comprimento, as condições de estringência muito baixa a muito alta são definidas como pré-hibridização e hibridização a 42°C em 5X SSPE, 0,3% SDS, 200 $\mu\text{g/ml}$ DNA de esperma de salmão desnaturado e cisalhado, e 25% formamida para estringências muito baixas e baixas, 35% formamida para
25 estringências média e média-alta, ou 50% formamida para estringências altas e muito altas, seguindo procedimentos Southern blotting padrões durante 12 a 24 horas, otimamente.

Para sondas longas de pelo menos 100 nucleotídeos em comprimento, o material veículo é finalmente lavado três vezes cada durante

15 minutos usando 2 X SSC, 02% SDS preferivelmente a 45°C (estringência muito baixa), mais preferivelmente a 50°C (estringência baixa), mais preferivelmente a 55°C (estringência média), mais preferivelmente a 60 °C (estringência média elevada), ainda mais preferivelmente a 65 °C (estringência alta), e o mais preferivelmente a 70 °C (estringência muito alta).

Para sondas curtas que tem cerca de 15 nucleotídeos a cerca de 70 nucleotídeos em comprimento, condições de estringência são definidas como pré-hibridização, hibridização e pós-hibridização de lavagem, a cerca de

5°C a cerca de 10 °C abaixo do T_m calculado, usando o cálculo de acordo com Bolton e McCarthy (1962, *Proceedings of the National Academy of Sciences USA* 48 1390) em 0,9 M NaCl, 0,09 M Tris-HCl pH 7,6, 6 mM EDTA, 0,5% NP- 40, 1X solução de Denhardt, 1 mM pirofosfato de sódio, 1 mM fosfato monobásico de sódio, 0,1 mM ATP, e 0,2 mg de RNA de levedura por ml seguindo os procedimentos de Southern blotting padrões durante 12 a 24 horas otimamente.

Para sondas curtas que tem cerca de 15 nucleotídeos a cerca de 70 nucleotídeos em comprimento, o material veículo é lavado uma vez em 6X SCC mais 0,1% SDS durante 15 minutos e duas vezes cada durante 15 min usando 6 X SSC a 5 °C a 10 °C abaixo do T_m calculado.

Polipeptídeos tendo atividade biológica e polinucleotídeos dos mesmos

Qualquer polipeptídeo que é fracamente secretado ou não secretado de todo pode ser usado nos métodos da presente invenção. O polipeptídeo pode ser qualquer polipeptídeo tendo uma atividade biológica de interesse. O polipeptídeo tendo atividade biológica pode ser nativo ou heterólogo (estranho) para a célula hospedeira fúngica de interesse. O termo "polipeptídeo heterólogo" é definido aqui como polipeptídeo que não é nativo para a célula hospedeira; ou um polipeptídeo nativo em que modificações estruturais foram feitas para alterar o polipeptídeo nativo.

Em um aspecto preferido, o polipeptídeo é uma anticorpo, antígeno, peptídeo antimicrobiano, enzima, fator de crescimento, hormônio, imunodilatador, neurotransmissor, receptor, proteína repórter, ou proteína estrutural

5 Em outro aspecto preferido, o polipeptídeo é uma albumina, colágeno, tropoelastina, elastina, ou gelatina. Em um aspecto mais preferido, o polipeptídeo é uma oxidoreductase, transferase, hidrolase, liase, isomerase, ou ligase.

10 Em um aspecto ainda mais preferido, o polipeptídeo é uma alfa-amilase, alfa- 1,3-glucanase, alfa-galactosidase, alfa-glucosidase, alfa-1,6-mannosidase, aminopeptidase, amilase, arabinase, beta-agarase beta-amilase, beta-1,3-glucanase beta-1,6-glucanase, beta-galactosidase, beta-glucosidase, beta-mannosidase carbohidrase, carboxipeptidase, catalase, celulase, chitinase, quitosanase, cutinase ciclodextrano glicosiltransferase, 15 deoxiribonuclease, dextranase, endo-1,4-beta- galactanase, endo-1,6-beta-galactanase, esterase, fucosidase, glucoamilase, glucocerebrosidase, hialuronidase, inulinase, invertase, laccase, levanase, liqueninase, lipase, lisozime, mannosidase, mutanase, oxidase, enzima pectinolítica, peroxidase, fosfolipase, fitase, poligalacturonase, polifenoloxidase, enzima proteolítica, 20 rhamnogalacturonase, rhamnosidase, ribonuclease, trehalase, transglutaminase, transglicosilase, uroquinase, ou xilanase.

Em ainda outro aspecto mais preferido, o polipeptídeo é uma enzima celulolítica ou celulase. Exemplos de enzimas celulolíticas incluem, mas não são limitados a endoglucanases, celobiohidrolases, e beta- 25 glucosidases. Outras proteínas que auxiliam a ação da enzima celulolítica são polipeptídeos tendo atividade melhora dora celulolítica (ver, por exemplo, WO 2005/074647, WO 2005/074656, e pedido US publicado 2007/0077630). Em um aspecto o mais preferido, o polipeptídeo é uma endoglucanase. Em outro aspecto o mais preferido, o polipeptídeo é uma celobiohidrolase. Em

outro aspecto o mais preferido, o polipeptídeo é uma beta-glucosidase. Em outro aspecto o mais preferido, o polipeptídeo é um polipeptídeo tendo atividade melhoradora celulolítica.

Em um outro aspecto o mais preferido, o polipeptídeo é uma hemicelulase. Hemicelulases podem ser colocadas em três categorias gerais: as enzimas de endo-atuação que atacam as ligações internas dentro da cadeia de polissacarídeo, as enzima de exo-atuação que atuam sucessivamente a partir ou da extremidade de redução ou não redução de cadeia de polissacarídeo, e as enzimas acessórias, acetilesterases e esterases que hidrolisam ligações de lignina glicosídeos, como esterase de ácido cumárico e esterase de ácido ferúlico (Wong, K. K. Y., Tan, L. U. L., e Saddler, J N, 1988, Multiplicity of β -1,4-xilanase in microorganisms: Functions and applications, *Microbiol Rev* 52 305-317, Tenkanen, M., e Poutanen, K, 1992, Significance of esterases in the degradation of xilans in *Xilans* and *Xylanases*, Visser, J., Beldman, G, Kuster-van Someren, M.A., e Voragen, A. G. J., eds, Elsevier, New York, NY, 203-212, Coughlan, M. P., e Hazlewood, G P, 1993, *Hemicellulose and hemicellulases*, Portland, London, UK, Brigham, J.S., Adney, W.S., e Mimmel, M. E., 1996, Hemicellulases Diversity and applications, em *Handbook on Bioethanol Production and Utilization*, Wyman, C E, ed, Taylor & Francis, Washington, DC, 119-141)

Hemicelulases incluem, mas não são limitadas a xilanases, arabinofuranosidasas, acetil xilano esterase, glucuronidasas, endo-galactanase, mananases, endo- ou exo- arabinases, exo-galactanases, e misturas das mesmas. Exemplos de hemicelulases de endo-atuação e enzimas auxiliares incluem, mas não são limitados a endoarabinanase, endoarabinogalactanase, endoglucanase, endomannanase, endoxilanase, e feraxano endoxilanase. Exemplos de hemicelulases de exo-atuação e enzimas auxiliares incluem, mas não são limitados a, alfa-L-arabinosidase, beta-L-arabinosidase, alfa-1,2-L- fucosidase, alfa-D-galactosidase, beta-D-

galactosidase, beta-D-glucosidase, beta-D- glucuronidase, beta-D-mannosidase, beta-D-xilosidase, exoglucosidase, exocelobiohidrolase, exomanobiohidrolase, exomananase, exoxilanase, xilano alfa-glucuronidase, e coniferina beta-glucosidase. Exemplos de esterases incluem, mas não são
 5 limitados a, acetil esterases (acetilgalactano esterase, acetilmanano esterase, e acetilxilano esterase) e aril esterases (esterase de ácido cumárico e esterase de ácido ferúlico).

Em outro aspecto o mais preferido, o polipeptídeo é uma xilanase. Em outro aspecto o mais preferido, o polipeptídeo é uma
 10 xiloglucanase. Em outro aspecto o mais preferido, o polipeptídeo é uma arabinofuranosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma glucuronidase, por exemplo, alfa-glucuronidase. Em outro aspecto o mais preferido, o polipeptídeo é uma endo-galactanase. Em outro aspecto o mais preferido, o polipeptídeo é uma mannanase. Em outro aspecto o mais
 15 preferido, o polipeptídeo é uma endo-arabinase. Em outro aspecto o mais preferido, o polipeptídeo é uma exo-arabinase. Em outro aspecto o mais preferido, o polipeptídeo é uma endoarabinanase. Em outro aspecto o mais preferido, o polipeptídeo é uma endoarabinogalactanase. Em outro aspecto o mais preferido, o polipeptídeo é uma endoglucanase. Em outro aspecto o mais
 20 preferido, o polipeptídeo é uma endomannanase. Em outro aspecto o mais preferido, o polipeptídeo é uma endoxilanase. Em outro aspecto o mais preferido, o polipeptídeo é uma feraxuma endoxilanase. Em outro aspecto o mais preferido, o polipeptídeo é uma alfa-L- arabinosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma beta-L- arabinosidase. Em outro
 25 aspecto o mais preferido, o polipeptídeo é uma beta-glucanase. Em outro aspecto o mais preferido, o polipeptídeo é uma alfa- 1,2-L-fucosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma alfa-D-galactosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma beta-D-galactosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma beta-

D-glucosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma beta-D-glucuronidase. Em outro aspecto o mais preferido, o polipeptídeo é uma beta-D-mannosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma beta-D-xilosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma exoglucosidase. Em outro aspecto o mais preferido, o polipeptídeo é uma exomannobiohidrolase. Em outro aspecto o mais preferido, o polipeptídeo é uma exomananase. Em outro aspecto o mais preferido, o polipeptídeo é uma exoxilanase. Em outro aspecto o mais preferido, o polipeptídeo é uma xilano alfa-glucuronidase. Em outro aspecto o mais preferido, o polipeptídeo é uma coniferina beta-glucosidase.

Em um outro aspecto o mais preferido, o polipeptídeo é uma esterase. Como usado aqui, uma "esterase" também conhecida como hidrolase de éster carboxílico, refere-se a enzimas atuando em ligações éster, e inclui enzimas classificadas em Hidrolases de éster carboxílico EC 3.1.1 de acordo com Nomenclatura de enzimas (*Enzyme Nomenclature*, 1992, Academic Press, San Diego, California, com Suplemento 1 (1993), Suplemento 2 (1994), Suplemento 3 (1995), Suplemento 4 (1997) e Suplemento 5, em *Eur. J. Biochem* 223 1-5, 1994, *Eur. J. Biochem* 232 1-6, 1995, *Eur. J. Biochem* 237 1-5, 1996, *Eur. J. Biochem* 250 1-6, 1997, e *Eur. J. Biochem* 264 610-650, 1999, respectivamente). Os exemplos não limitativos de esterases incluem arilesterase, triacilglicerol lipase, acetilesterase, acetilcolinaesterase, colinaesterase, tropinaesterase, pectinaesterase, esterol esterase, clorofilase, L-arabinonolactonase gluconolactonase, uronolactonase, tanase, retinil-palmitato esterase, hidroxibutirato- dímero hidrolase, acilglicerol lipase, 3-oxoadipato enol-lactonase, 1,4-lactonase, galactolipase, 4-piridoxolactonase, acilcarnitina hidrolase, aminoacil-tRNA hidrolase, D-arabinonolactonase, 6-fosfogluconolactonase, fosfolipase A1, 6-acetilglicose deacetilase, lipoproteína lipase, dihidrocoumarina lipase, limonina-D-anel-lactonase, esteróide-lactonase, triacetato-lactonase, actinomicina lactonase, orselinato-

depsídeo hidrolase, cefalosporina-C deacetilase, clorogenato hidrolase, alfa-aminoácido esterase, 4-metiloxaloacetato esterase, carboximetilno butenolidase, deoxilimonate A-anel-lactonase, 2-acetil-1-alquilglicerofosfocolina esterase, fusarinina- C- ornitinaesterase, sinapina esterase, cera - éster hidrolase, forbol-diéster hidrolase, fosfatidilinositol deacilase, sialato O-acetilesterase, acetoxibutinilbitiofeno deacetilase, acetilsalicilato deacetilase, metilumbelliferil-acetato deacetilase, 2-pirona-4,6-dicarboxilato lactonase, N- acetilgalactosaminoglicano deacetilase, esterase de hormônio juvenil, bis(2- etilhexil)ftalato esterase, proteína-glutamato metilesterase, 11-cis retinil-palmitato hidrolase, todo- trans-retinil-palmitato hidrolase, L-rhamnono-1,4-lactonase, 5-(3,4- diacetoxibut-1-inil)-2,2'-bitiofeno deacetilase, ácido graxo - étil- éster sintase, xilono-1,4-lactonase, N-acetilglucosaminilfosfatidilinositol deacetilase, cetraxato benzilesterase, e acetilalquilglicerol acetilhidrolase Esterases que podem ser usadas para a bioconversão de celulose incluem acetil esterases como acetilgalactano esterase, acetilmanano esterase, e acetilxilano esterase, e esterases que hidrolisam ligações lignina glicosídeos, como esterase de ácido cumárico e esterase de ácido ferúlico.

Em outro aspecto o mais preferido, o polipeptídeo é uma acetil esterase. Em outro aspecto o mais preferido, o polipeptídeo é uma acetilgalactano esterase codificando, o polipeptídeo é uma acetilmannano esterase. Em outro aspecto o mais preferido, o polipeptídeo é uma acetil xilano esterase. Em outro aspecto o mais preferido, o polipeptídeo é uma aril esterase. Em outro aspecto o mais preferido, o polipeptídeo é uma esterase de ácido cumárico. Em outro aspecto o mais preferido, o polipeptídeo é uma esterase de ácido ferúlico.

Em um outro aspecto o mais preferido, o polipeptídeo é uma lipase. Em um outro aspecto o mais preferido, o polipeptídeo é uma fosfolipase, por exemplo, fosfolipase A1, fosfolipase A2, fosfolipase C,

fosfolipase C, ou fosfolipase D.

Em um outro aspecto o mais preferido, o polipeptídeo é uma cutinase.

5 Em outro aspecto o mais preferido, o polipeptídeo é uma glicose isomerase. Em outro aspecto o mais preferido, o polipeptídeo é uma xilose isomerase.

Em um outro aspecto o mais preferido, o polipeptídeo é uma enzima proteolítica. Proteases são bem conhecidas na arte e se referem a enzimas que catalisam a clivagem de ligações de peptídeos.

10 Em outro aspecto o mais preferido, o polipeptídeo é uma serina protease. Em outro aspecto o mais preferido, o polipeptídeo é uma metaloprotease. Em outro aspecto o mais preferido, o polipeptídeo é uma tiol protease.

15 Em um outro aspecto o mais preferido, o polipeptídeo é uma peptidase. Em outro aspecto o mais preferido, o polipeptídeo é uma aminopeptidase, por exemplo dipeptidilaminopeptidase ou tripeptidilaminopeptidase. Em outro aspecto o mais preferido, o polipeptídeo é uma carboxipeptidase.

20 Em um outro aspecto o mais preferido, o polipeptídeo é uma laccase.

Em um outro aspecto o mais preferido, o polipeptídeo é uma peroxidase.

25 Em um outro aspecto o mais preferido, o polipeptídeo é uma enzima degradando amido. Em outro aspecto o mais preferido, o polipeptídeo é uma alfa-amilase. Em outro aspecto o mais preferido, o polipeptídeo é uma amiloglucosidase. Em outro aspecto o mais preferido, o polipeptídeo é pululanase. Em outro aspecto o mais preferido, o polipeptídeo é uma enzima desramificadora. Em outro aspecto o mais preferido, o polipeptídeo é uma ciclodextrina glicosiltransferase.

Um polinucleotídeo codificando um domínio catalítico, polipeptídeo maduro, ou polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, ou uma porção da mesma, podem ser obtidos a partir de qualquer organismo.

5 Um polinucleotídeo codificando um polipeptídeo tendo atividade biológica pode ser obtido a partir de um gene codificando um polipeptídeo bacteriano. Por exemplo, o polipeptídeo pode ser um polipeptídeo bacteriano Gram positivo incluindo, mas não limitado a, um polipeptídeo de *Bacillus*, *Streptococcus*, *Streptomyces*, *Staphylococcus*,
10 *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus*, ou *Oceanobaccillus*, ou um polipeptídeo bacteriano Gram negativo, incluindo, mas não limitado a, polipeptídeo de *E. coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*, *Flavobacterium*, *Fusobacterium*, *Llyobacter* *Neisseria*, ou *Ureaplasma*.

15 Em um aspecto preferido, um polinucleotídeo codificando um polipeptídeo tendo atividade biológica pode ser obtido a partir de um gene codificando um polipeptídeo de *Bacillus alkalophilus*, *Bacillus amiloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus tautus*, *Bacillus lentus*, *Bacillus*
20 *hcheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, ou *Bacillus thuringiensis*.

Em outro aspecto preferido, um polinucleotídeo codificando um polipeptídeo tendo atividade biológica pode ser obtido a partir de um gene codificando um polipeptídeo de *Streptococcus equisimilis*, *Streptococcus*
25 *pyogenes*, *Streptococcus ubens*, ou *Streptococcus equi subsp Zooepidemicus*.

Em outro aspecto preferido, um polinucleotídeo codificando um polipeptídeo tendo atividade biológica pode ser obtido a partir de um gene codificando um polipeptídeo de *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, ou *Streptomyces*

lividans.

Um polinucleotídeo codificando um polipeptídeo tendo atividade biológica também pode ser obtido a partir de um gene codificando um polipeptídeo fúngico, e mais preferivelmente um polipeptídeo de levedura
 5 como um polipeptídeo de *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schtzosaccharomyces*, ou *Yarrowia*, ou mais preferivelmente um polipeptídeo de fungo filamentoso como um polipeptídeo de *Acremonium*, *Agaricus*, *Atternana*, *Aspergillus*, *Aureobasidium*, *Botryospaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobotus*, *Coprinopsis*,
 10 *Coptotermes*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptospaena*, *Magnaporthe*, *Melanocarpus*, *Menpilus*, *Mucor*, *Mycehophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania* *Pseudotnchonympha*,
 15 *Rhizomucor*, *Schizophyllum*, *Scytalidium*, *Talaromyces*, *Thermoaseus*, *Thielavia*, *Totypodadium*, *Trichoderma*, *Trichophaea*, *Verticialium*, *Volvariella*, ou *Xilaria*.

Em um aspecto preferido, um polinucleotídeo codificando um polipeptídeo tendo atividade biológica pode ser obtido a partir de um gene
 20 codificando um polipeptídeo de *Saccharomyces cartsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasu*, *Saccharomyces kluyver*, *Saccharomyces norbensis*, ou *Saccharomyces oviformis*

Em outro aspecto preferido, um polinucleotídeo codificando
 25 um polipeptídeo tendo atividade endoglucanase pode ser obtido a partir de um gene codificando um polipeptídeo de *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus* *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosponum keratinophilum*, *Chrysosponum lucknowense*,

Chrysosponum tmpicum, *Chrysosponum merdarium*, *Chrysosporium mops*,
Chrysosponum pannicola, *Chrysosponum queenslandicum*, *Chrysosponum*
zonatum, *Cladorrhinum foecundissimum*, *Fusarium bactndioides*, *Fusarium*
cerealis, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium*
5 *graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium*
negundi, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*,
Fusarium sambuanum, *Fusarium sarcochroum*, *Fusarium sporotnchioides*,
Fusarium sulphureum, *Fusarium torulosum*, *Fusarium trichothecioides*,
Fusarium venenatum, *Humicola grisea*, *Humicola insolens*, *Humicola*
10 *lanuginosa*, *Irpex lacteus*, *Mucor miehei*, *Myceliophthora thermophila*,
Neurospora crassa, *Penicillmm funiculosum*, *Pemallium purpurogenum*,
Phanerochaete chrysosponum, *Thielavia achromatica*, *Thielavia albomyces*,
Thielavia albopilosa, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia*
microspore, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia spededomum*,
15 *Thielavia setosa*, *Thielavia subthermophila*, *Thielavia terrestris*, *Thielavia*
vanospora, *Thielavia wareingii*, *Trichoderma harzianum*, *Trichoderma*
kontngii, *Trichoderma longibrachiatum* *Trichoderma reesei*, ou *Trichoderma*
viride.

Em um aspecto particularmente preferido, o polipeptídeo
 tendo atividade biológica é uma beta-glucosidase. Exemplos de beta-
 20 glucosidases que podem ser usadas como fontes para os polinucleotídeos nos
 métodos da presente invenção, incluem, mas não são limitados a, beta-
 glucosidase de *Aspergillus oryzae* (WO 02/095014, WO 04/099228), beta-
 glucosidase de *Aspergillus aculeatus* (Kawaguchi et al, 1996, *Gene* 173 287-
 25 288), *Aspergillus avenaceus* (GENBANK™ número de acesso AY943971),
Aspergillus fumigatus (GENBANK™ número de acesso XM745234),
Aspergillus kawachii(GENBANK™ número de acesso AB003470),
Aspergillus niger (GENBANK™ AJ132386), *Magnaporthe grisea*
 (GENBANK™ número de acesso AY849670), *Phanerochaete chrysosponum*

(GENBANK™ número de acesso AB253327), *Talaromyces emersonii* (GENBANK™ número de acesso AY072918), e *Trichoderma reesei* (GENBANK™ números de acesso U09580, AB003110, AY281374, AY281375, AY281377, AY281378, e AY281379). Variantes de beta-glucosidases também podem ser usadas como fontes para os polinucleotídeos como os descritos em WO 04/099228.

Em um aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase Família 1.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase Família 3.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase de *Aspergillus oryzae*. Em um aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene beta-glucosidase de *Aspergillus oryzae* compreendendo SEQ ID NO 23 que codifica o polipeptídeo de SEQ ID NO 24 ou um gene mutante de beta-glucosidase de *Aspergillus oryzae* compreendendo SEQ ID NO 25 que codifica o polipeptídeo de SEQ ID NO 26.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase de *Aspergillus fumigatus*. Em um aspecto o mais

preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase de *Aspergillus fumigatus* compreendendo SEQ ID NO 35 27 que
5 codifica o polipeptídeo de SEQ ID NO 28.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene beta-glucosidase de *Penicillium brasilianum* cepa IBT 20888. Em
10 um aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta- glucosidase de *Penicillium brasilianum* cepa IBT 20888 compreendendo SEQ ID NO 29 que codifica o polipeptídeo de SEQ ID NO
15 30.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase de *Trichoderma reesei* cepa No QM9414. Em
20 outro aspecto o mais preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene de beta-glucosidase de *Trichoderma reesei* cepa No QM9414 compreendendo SEQ ID NO 31 que codifica o polipeptídeo de SEQ ID NO
25 32 (GENBANK™ número de acesso U09580). Em outro aspecto preferido, a beta-glucosidase é naturalmente secretada. Em outro aspecto preferido, a beta-glucosidase não é naturalmente secretada.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou

uma porção da mesma, é codificado por um polinucleotídeo obtido a partir de um gene codificando um polipeptídeo homólogo compreendendo uma sequência de aminoácidos que tem um grau de identidade para as sequências de aminoácidos do polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico de SEQ ID NO 24, SEQ ID NO 26, SEQ ID NO 28, SEQ ID NO 30, ou SEQ ID NO 32 de pelo menos 60%, preferivelmente pelo menos 65%, mais preferivelmente pelo menos 70%, mais preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, mais preferivelmente pelo menos 85%, ainda mais preferivelmente pelo menos 90%, o mais preferivelmente pelo menos 95%, e ainda o mais preferivelmente 96%, 97%, 98%, ou 99%, que tem atividade endoglucanase. Em um aspecto preferido, o polipeptídeo homólogo tem uma sequência de aminoácidos que difere por dez aminoácidos, preferivelmente por cinco aminoácidos, mais preferivelmente por quatro aminoácidos, ainda mais preferivelmente por três aminoácidos, o mais preferivelmente por dois aminoácidos, e ainda o mais preferivelmente por um aminoácido a partir do polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico de SEQ ID NO 24, SEQ ID NO 26, SEQ ID NO 28, SEQ ID NO 30, ou SEQ ID NO 32.

Em outro aspecto preferido, o polipeptídeo de comprimento completo, polipeptídeo maduro, ou domínio catalítico da beta-glucosidase, ou uma porção da mesma, é codificado por um polinucleotídeo que hibridiza sob condições de estringência muito baixa, preferivelmente condições de estringência baixa, mais preferivelmente condições de estringência média, mais preferivelmente condições de estringência média-alta, ainda mais preferivelmente condições de estringência alta, e o mais preferivelmente condições de estringência muito alta com (i) SEQ ID NO 23, SEQ ID NO 25, SEQ ID NO 27, SEQ ID NO 29, ou SEQ ID NO 31, (ii) a sequência de cDNA contida em SEQ ID NO 23, SEQ ID NO 25, SEQ ID NO 27, SEQ ID NO 29,

ou SEQ ID NO 31, (ii) a subsequência de (i) ou (ii) ou (iv) um filamento complementar de (i), (ii) ou (iii), (J.Sambrook, E.F. Fritsch, e T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2d ed., Cold Spring Harbor, New York). A subsequência de SEQ ID NO 23, SEQ ID NO 25, SEQ ID NO 27, SEQ ID NO 29, ou SEQ ID NO 31 contém pelo menos 100 nucleotídeos contíguos ou preferivelmente pelo menos 200 nucleotídeos contíguos.

A sequência de nucleotídeos de SEQ ID NO 23, SEQ ID NO 25, SEQ ID NO 27 SEQ ID NO 29, ou SEQ ID NO 31, ou a subsequência da mesma, assim como a sequência de aminoácidos de SEQ ID NO 24, SEQ ID NO 26, SEQ ID NO 28, SEQ ID NO 30, ou SEQ ID NO 32, ou um fragmento da mesma, pode ser usada para projetar uma sonda de ácido nucleico para identificar e clonar polipeptídeos codificando DNA tendo atividade beta-glucosidase a partir de cepas de diferentes gêneros ou espécies acordo com métodos bem conhecidos na arte. Em particular, estas sondas podem ser usadas para hibridização com o cDNA ou genômico do gênero ou espécie de interesse, seguindo os procedimentos de Southern blotting padrões, ou a fim de identificar e isolar o gene correspondente ai. Estas sondas são consideravelmente mais curtas do que a sequência completa, mas devem ser de pelo menos 14, preferivelmente pelo menos 25, mais preferivelmente pelo menos 35, e o mais preferivelmente pelo menos 70 nucleotídeos em comprimento. No entanto, prefere-se que a sonda de ácido nucleico tenha pelo menos 100 nucleotídeos em comprimento. Por exemplo, a sonda de ácido nucleico pode ter pelo menos 200 nucleotídeos, preferivelmente pelo menos 300 nucleotídeos, mais preferivelmente pelo menos 400 nucleotídeos, ou o mais preferivelmente pelo menos 500 nucleotídeos em comprimento. Sondas ainda mais longas podem ser usadas, por exemplo, sondas de ácido nucleico que tem pelo menos 600 nucleotídeos, pelo menos preferivelmente pelo menos 700 nucleotídeos, mais preferivelmente pelo menos 800 nucleotídeos, ou o mais preferivelmente pelo menos 900 nucleotídeos em comprimento.

Ambas as sondas de DNA e RNA podem ser usadas. As sondas são tipicamente rotuladas para detectar o gene correspondente (por exemplo, com ^{32}P , ^3H , ^{35}S , biotina, ou avidina). Estas sondas são englobadas pela presente invenção.

5 **Peptídeos sinal.**

O peptídeo sinal pode ser qualquer peptídeo sinal apropriado reconhecido por uma célula hospedeira para secreção extracelular de uma proteína de fusão da presente invenção. A sequência de sinal é preferivelmente a que está naturalmente associada com o componente endoglucanase de uma proteína de fusão a ser expressada.

A extremidade 5' da sequência de codificação da sequência de nucleotídeos codificando um polipeptídeo que contém inerentemente uma região de codificação de peptídeo naturalmente ligada em matriz de leitura de tradução com o segmento da região de codificação que codifica o polipeptídeo secretado. Alternativamente, a extremidade 5' da sequência de codificação pode conter uma região de codificação de peptídeo sinal que é estranha para a sequência de codificação. A região de codificação de peptídeo sinal estranho pode ser requerida onde a sequência de codificação não contém naturalmente uma região de codificação de peptídeo sinal. Alternativamente, a região de codificação de peptídeo sinal estranho pode simplesmente substituir a região de codificação de peptídeo sinal natural a fim de melhorar a secreção do polipeptídeo. No entanto, qualquer região de codificação de peptídeo sinal que dirige a proteína de fusão expressada na via secretória da célula hospedeira de escolha, isto é, secretada em um meio de cultura, pode ser usada na presente invenção.

As sequências de codificação de peptídeo sinal eficazes para células hospedeiras de fungos filamentosos são a sequência de codificação de peptídeo sinal obtidas a partir dos genes para TAKA amilase de *Aspergillus oryzae*, amilase neutral de *Aspergillus niger*, glucoamilase de *Aspergillus*

niger, aspártica proteinase de *Rhizomucor miehei*, celulase de *Humicola insolens*, endoglucanase V de *Humicola insolens*, e lipase de *Humicola lanuginosa*, celobiohidrolase I de *Trichoderma reesei*, celobiohidrolase II de *Trichoderma reesei*, endoglucanase I de *Trichoderma reesei*, endoglucanase II de *Trichoderma reesei*, endoglucanase III *Trichoderma reesei*, endoglucanase IV *Trichoderma reesei*, endoglucanase V *Trichoderma reesei*, xilanase I de *Trichoderma reesei*, xilanase II de *Trichoderma reesei*, e beta-xilosidase de *Trichoderma reesei*.

As sequências de codificação de peptídeo sinal utilizáveis para células hospedeiras de levedura são obtidas dos genes para alfa-fator de *Saccharomyces cerevisiae* e invertase de *Saccharomyces cerevisiae*. Outra sequência de codificação de peptídeo sinal utilizável é descrita por Romanos et al, 1992, supra.

Em um aspecto preferido, a sequência de codificação de peptídeo sinal é obtida a partir de um gene que codifica uma endoglucanase. Em um aspecto mais preferido, a sequência de codificação de peptídeo sinal é obtido a partir de um gene que codifica uma endoglucanase VI. Em um aspecto ainda mais preferido, a sequência de codificação de peptídeo sinal é obtido a partir de um gene de *Humicola insolens* que codifica uma endoglucanase V. Em um aspecto ainda mais preferido, a sequência de codificação de peptídeo sinal codifica aminoácidos 1 a 21 de SEQ ID NO 2. Em outro aspecto o mais preferido, a sequência de codificação de peptídeo sinal é de nucleotídeos 1 a 63 de SEQ ID NO 1.

A proteína de fusão pode ainda compreender um segundo peptídeo sinal que é associado com o componente de beta-glucosidase de uma proteína de fusão. A sequência de codificação de peptídeo sinal pode ser uma sequência de codificação de peptídeo sinal que está naturalmente associada com a sequência de codificação do polipeptídeo tendo atividade biológica ou pode ser uma sequência de codificação de peptídeo sinal diferente como uma

das descritas acima.

Em um aspecto preferido, a sequência de codificação de peptídeo sinal é uma sequência naturalmente associada com a sequência de codificação de beta-glucosidase.

5 **Ligadores**

Como mencionado supra, muitas endoglucanases tem uma estrutura de múltiplos domínios consistindo de um domínio catalítico separado de um ou mais módulos de ligação de carboidratos por um peptídeo (s) ligador (es). Nos métodos da presente invenção, as construções de proteína de fusão pode ainda compreender uma sequência de codificação de ligador localizada 3' à sequência compreendendo o domínio catalítico de endoglucanase e 5' para a sequência compreendendo o domínio catalítico do polipeptídeo tendo atividade biológica.

O ligador pode ser obtido do mesmo gene que o domínio catalítico da endoglucanase ou de um gene de endoglucanase diferente. Por outro lado, o ligador pode ser de origem sintética.

Os exemplos de ligadores que podem ser usados nos métodos da presente invenção incluem, mas não limitados a ligadores obtidos dos genes de genes de celobiohidrolase I de *Trichoderma reesei* (Srisodsuk et al, 1993, *Journal of Biological Chemistry* 268 20765-20761), celobiohidrolase de *Hypocrea jecoonna* (antes *Trichoderma reesei*) Cel7A (Mulakala et al, 2005, *Proteins* 60 598-605), endoglucanase V de *Humicola insolens*, e endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7C.

Em um aspecto preferido, o ligador é obtido a partir de um gene de endoglucanase de *Humicola insolens*. Em outro aspecto preferido, o ligador é obtido a partir de um gene de endoglucanase de *Trichoderma reesei*. Em um aspecto mais preferido, o ligador é obtido a partir de um gene de endoglucanase V de *Humicola insolens* (eg5).

Em outro aspecto preferido, o ligador é obtido a partir de um

gene de endoglucanase de *Thielavia terrestris*. Em um outro aspecto o mais preferido, o ligador é obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7C.

5 Em um aspecto preferido, o ligador tem pelo menos 5 resíduos de aminoácido. Em um aspecto mais preferido, o ligador tem pelo menos 15 resíduos de aminoácido. Em um aspecto o mais preferido, o ligador tem pelo menos 25 resíduos de aminoácido.

10 Em um aspecto preferido, o ligador tem entre cerca de 5 a cerca de 60 resíduos de aminoácido. Em um aspecto mais preferido, o ligador tem entre cerca de 15 a cerca de 50 resíduos de aminoácido. Em um aspecto o mais preferido, o ligador tem entre cerca de 25 a cerca de 45 resíduos de aminoácido.

Módulos de ligação de carboidratos

15 Módulos de ligação de carboidratos (CBMs) são definidos como sequências de aminoácidos contíguas com uma duplicação discreta tendo atividade de ligação de carboidratos, e são comumente encontrados dentro de enzimas ativas para carboidratos. Vários tipos de CBMS foram descritos e a maior parte liga a polissacarídeos insolúveis (ver Boraston et al, 2004, *BioChem J.* 382 769-781)

20 Módulos de ligação de carboidratos foram caracterizados os quais mediam a interação com, por exemplo, celulose cristalina, celulose não cristalina, quitina, beta-1,3-glucanos e beta-1,3-1,4-glucanos de ligação mista, xilano, mannano, galactano e amido. Módulos de ligação de carboidratos ocorrem com frequência em celulasas de múltiplos domínios. Apesar de
25 alguns CBMS conferirem ligação específica a um subconjunto de estruturas de carboidratos, outros são mais gerais em sua capacidade de se associar com vários polissacarídeos. CBMs que conferem ligação a celulose são às vezes referidos como domínios de ligação de celulose, ou CBDs (Boraston et al, 2004, *Biochem J.* 382 769-781). CBMs são agrupados por similaridade com

aminoácido, atualmente 48 famílias de CBM são descritas.

As glicosídeo hidrolases podem compreender mais de um domínio catalítico e um, dois, três ou mais CBMs, e opcionalmente ainda compreender uma ou mais regiões de sequência de aminoácido de polipeptídeo ligando os CBM(s) com o(s) domínio(s) catalítico(s), uma região do último tipo geralmente sendo denotada um "ligador". Os exemplos de enzimas hidrolíticas compreendendo um CBM são e celulasas, xilanasas, mannanases, arabinofuranosidasas acetilesterases e quitinasas (ver P. Tomme et al, *Cellulose -Binding Domains - Classification and Properties in Enzymatic Degradation of Insoluble Carbohydrates*, John N. Saddler e Michael H. Penner (Eds), ACS Symposium Series, No 618, 1996).

Um CBM pode estar localizado no término N ou C ou em uma posição interna de uma proteína ou polipeptídeo. A região de um polipeptídeo ou proteína que constitui um CBM tipicamente consiste de mais do que cerca de 30 e menos do que cerca de 250 resíduos de aminoácidos. Por exemplo os CBMs listados e classificados em Família I consistem de 33-37 resíduos de aminoácido, os listados e classificados em Família 2a consistem de 95-108 resíduos de aminoácido, e os listados e classificados em Família 6 consistem de 85-92 resíduos de aminoácido. Assim, o peso molecular de uma sequência de aminoácidos constituindo um CBM estará tipicamente na faixa de cerca de 4 kDa a cerca de 40 kDa, e geralmente abaixo de cerca de 35 kDa.

Nos métodos da presente invenção, qualquer CBM pode ser usado. O CBM pode ser naturalmente associado com a endoglucanase ou pode ser estranho para a endoglucanase.

Em um aspecto preferido, um CBM é obtido a partir de um gene de endoglucanase de *Trichoderma reesei* (EG). Em um aspecto mais preferido, um CBM é obtido a partir de um gene de endoglucanase EGI de *Trichoderma reesei*. Em um outro aspecto o mais preferido, um CBM é obtido a partir de um gene endoglucanase EGII de *Trichoderma reesei*. Em

um outro aspecto o mais preferido, um CBM é obtido a partir de um gene de endoglucanase EGV de *Trichoderma reesei*. Em outro aspecto preferido, um CBM é obtido a partir de um gene de celobiohidrolase (CBH) de *Trichoderma reesei* gene. Em outro aspecto preferido, um CBM é obtido a partir de um gene CBHI *Trichoderma reesei* (Tern et al, 1987, *Gene* 51 42-52, Linder e Teen, 1996, *Biochemistry* 93 12251-12255). Em outro aspecto preferido, um CBM é obtido a partir de um gene CBHII de *Trichoderma reesei*.

Em outro aspecto preferido, um CBM é obtido a partir de um gene de endoglucanase de *Thielavia terrestris*. Em um outro aspecto o mais preferido, um CBM é obtido a partir de um gene de endoglucanase de *Thielavia terrestris* NRRL 8126 CEL7C.

Sítio de clivagem

Nos métodos da presente invenção, as construções de proteína de fusão podem ainda compreender uma sequência de nucleotídeos codificando um sítio de clivagem. O sítio de clivagem está preferivelmente localizado entre a sequência compreendendo pelo menos o domínio catalítico de endoglucanase e a sequência compreendendo pelo menos o domínio catalítico do polipeptídeo tendo atividade biológica. Quando da secreção de uma proteína de fusão, o sítio é clivado liberando o polipeptídeo tendo atividade biológica de uma proteína de fusão.

Exemplos de sítios de clivagem incluem, mas não são limitados a, sítio Kex2 que codifica o dipeptídeo Lys-Arg (Martin et al, 2003, *J. Ind. Microbiol. Biotechnol* 3 568- 76, Svetina et al, 2000, *J Biotechnol* 76 245-251, Rasmussen-Wilson Et al, 1997, *Appl Environ Microbiol* 63 3488-3493, Ward et al, 1995, *Biotechnology* 13 498-503, e Contreras et al, 1991, *Biotechnology* 9 378-381), um sítio Me-(Glu ou Asp)-Gly-Arg, que é clivado por um Fator Xa protease após o resíduo arginina (Eaton et al, 1986, *Biochem* 25 505-512), um sítio Asp-Asp-Asp-Asp-Lys site, que é clivado por uma enteroquinase após a lisina (Colhns-Racie et al, 1995, *Biotechnology* (982-

987), um sítio His-Tyr-Glu ou sítio His-Tyr-Asp, que é clivado por Genenase I (Carter et al, 1989, *Proteins Structure, Function, and Genetics* 240-248), um sítio Leu-Val-Pro-Arg-Gly-Ser que é clivado por trombina após o Arg (Stevens, 2003, *Drug Discovery World* 4 35- 48J, um sítio Glu-Asn-Leu-Tyr-Phe-Gln-Gly, que é clivado por TEV protease após o Gln (Stevens, 2003, supra), e um sítio Leu- Glu-Val-Leu-Phe-Gln-Gly-Pro, que é clivado por forma geneticamente engenheirada de rinovírus humano 3C protease após o Gln (Stevens, 2003, supra)

Proteínas de fusão

10 Uma proteína de fusão tendo atividade biológica da presente invenção compreendendo um peptídeo sinal, pelo menos o domínio catalítico de uma endoglucanase ou uma porção da mesma, e pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica ou uma porção do mesmo, aumenta a secreção de uma proteína de fusão comparado com a
15 ausência de pelo menos o domínio catalítico da endoglucanase ou uma porção da mesma. Em cada um dos aspectos preferidos abaixo, os componentes de uma proteína de fusão são ligados em armação a partir do término N para o término C da proteína.

20 Em um aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, e um domínio catalítico de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, e um domínio catalítico de uma endoglucanase.

25 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, e um domínio catalítico de um polipeptídeo tendo atividade biológica

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade

biológica, e um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

5 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um domínio catalítico de uma endoglucanase.

10 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, e um polipeptídeo maduro de uma endoglucanase.

15 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, e um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

20 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, e um domínio catalítico de uma endoglucanase.

25 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, e um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, e um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a proteína de fusão compreende um polipeptídeo de comprimento completo de uma endoglucanase (peptídeo sinal e polipeptídeo maduro), e um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

5 Em outro aspecto preferido, a proteína de fusão compreende um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica e um polipeptídeo de comprimento completo de uma endoglucanase (peptídeo sinal e polipeptídeo maduro).

10 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, e um domínio catalítico de um polipeptídeo tendo atividade biológica.

15 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, e um domínio catalítico de uma endoglucanase.

20 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, e um domínio catalítico de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo maduro de uma endoglucanase.

25 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende

um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato e um domínio catalítico de uma endoglucanase.

5 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

10 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo maduro de uma endoglucanase.

15 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, e um domínio catalítico de uma endoglucanase.

20 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica.

25 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo de comprimento completo de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, e um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um segundo

peptídeo sinal, e um domínio catalítico de um polipeptídeo tendo atividade biológica.

5 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, um segundo peptídeo sinal, e um domínio catalítico de uma endoglucanase.

10 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um segundo peptídeo sinal, e um domínio catalítico de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, um segundo peptídeo sinal, e um polipeptídeo maduro de uma endoglucanase.

15 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um segundo peptídeo sinal, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

20 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um segundo peptídeo sinal, e um domínio catalítico de uma endoglucanase.

25 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um segundo peptídeo sinal, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um segundo peptídeo sinal, e um polipeptídeo maduro de

uma endoglucanase.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal; e um
5 domínio catalítico de um polipeptídeo tendo atividade biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um domínio catalítico de uma endoglucanase.

10 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um domínio catalítico de um polipeptídeo tendo atividade biológica.

15 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um polipeptídeo maduro de uma endoglucanase.

20 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um domínio catalítico de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica.

25 Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um domínio catalítico de uma endoglucanase.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, um polipeptídeo maduro de uma endoglucanase, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um polipeptídeo maduro de um polipeptídeo tendo atividade

biológica.

Em outro aspecto preferido, a proteína de fusão compreende um peptídeo sinal, e um polipeptídeo maduro de um polipeptídeo tendo atividade biológica, um ligador e/ou um módulo de ligação de carboidrato, um segundo peptídeo sinal, e um polipeptídeo maduro de uma endoglucanase.

Em outro aspecto preferido, para cada um dos aspectos preferidos acima, a proteína de fusão pode compreender alternativamente uma porção do domínio catalítico, o polipeptídeo maduro, ou o polipeptídeo de comprimento completo de uma endoglucanase ou um polipeptídeo tendo atividade biológica. A porção da endoglucanase pode ter ou não ter atividade de endoglucanase. Em um aspecto mais preferido, a porção da endoglucanase tem atividade endoglucanase.

Promotores

A região de promotor pode ser qualquer sequência de promotor apropriada reconhecida por uma célula hospedeira para expressão de uma proteína de fusão. A sequência de promotor contém sequências de controle transcripcional que mediam a expressão do polipeptídeo. O promotor pode ser qualquer sequência de nucleotídeo que mostra atividade transcripcional na célula hospedeira de escolha incluindo promotores mutantes, truncados, tandem, e híbridos e pode ser obtido de genes codificando polipeptídeos extracelulares ou intracelulares ou nativos ou estranhos (heterólogos) para a célula hospedeira. Os promotores exemplares incluem tanto promotores constitutivos como indutíveis.

Os exemplos de promotores apropriados para dirigir a transcrição das construções de proteína de fusão da presente invenção, especialmente em uma célula hospedeira bacteriana, são os promotores obtidos de operon lac de *E. coli*, gene agarase de *Streptomyces coelicolor* (dagA), gene levansucrase de *Bacillus subtilis* (sacB), gene alfa-amilase de *Bacillus licheniformis* (amyL), gene maltogênico amilase de *Bacillus*

stearothermophilus (amyM), gene alfa-amilase de *Bacillus amiloliquefaciens* (amyQ), gene penicillinase de *Bacillus licheniformis* (penP), genes xylA e xylB de *Bacillus subtilis*, e gene beta-lactamase procariótico (Villa-Kamaroff et al, 1978, *Proceedings of the National Academy of Sciences USA* 75 3727-3731), assim como um *promotor tac* (DeBoer et al, 1983, *Proceedings of the National Academy of Sciences USA* 80 21-25) outros promotores são descritos em "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242 74-94, e em Sambrook et al, 1989, supra.

Exemplos de promotores apropriadas para dirigir a transcrição de construções de fusão da presente invenção em uma célula hospedeira de fungo filamentosos são promotores obtidos a partir dos genes para TAKA amilase de *Aspergillus oryzae*, proteinase aspártica de *Rhizomucor miehei*, alfa-amilase neutral de *Aspergillus niger*, alfa-amilase estável em ácido de *Aspergillus niger* acid stable, glucoamilase de *Aspergillus niger* ou *Aspergillus awamori* (glaA), lipase de *Rhizomucor miehei*, protease alcalina de *Aspergillus oryzae*, triose fosfato isomerase de *Aspergillus oryzae*, acetamidase de *Aspergillus nidulans*, amiloglucosidase de *Fusarium venenatum* (WO 00/56900), *Fusarium venenatum* Dana (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), protease tipo tripsina de *Fusarium oxysporum* (WO 96/00787), beta-glucosidase de *Trichoderma reesei*, celobiohidrolase I de *Trichoderma reesei*, celobiohidrolase II de *Trichoderma reesei*, endoglucanase I de *Trichoderma reesei*, endoglucanase II de *Trichoderma reesei*, endoglucanase III de *Trichoderma reesei*, endoglucanase IV de *Trichoderma reesei*, endoglucanase V de *Trichoderma reesei*, xilanase I de *Trichoderma reesei*, xilanase II de *Trichoderma reesei*, beta-xilosidase de *Trichoderma reesei*, beta-tubulina de *Coprinus cinereus* e swollenina de *Trichoderma reesei*, assim como promotor NA2-tpi (um híbrido dos promotores dos genes para alfa-amilase neutral *Aspergillus niger* e triose fosfato isomerase *Aspergillus oryzae*), e os promotores mutantes,

truncados e híbridos dos mesmos.

Em hospedeiro de levedura, promotores apropriados são obtidos a partir de genes para enolase de *Saccharomyces cerevisiae* (ENO-1), galactosquinase de *Saccharomyces cerevisiae* (GAL1), álcool dehidrogenase/gliceraldeído-3- fosfato dehidrogenase *Saccharomyces cerevisiae* (ADH1, ADH2/GAP), triose fosfato isomerase *Saccharomyces cerevisiae* (TPI), metalotionina *Saccharomyces cerevisiae* (CUP1), e fosfoglicerato quinase de *Saccharomyces cerevisiae*. Outros promotores utilizáveis para as células hospedeiras de levedura são descritos por Romanos et al, 1992, *Yeast* 8 423-488.

Em um aspecto preferido, o promotor é um promotor de celobiohidrolase. Em um aspecto mais preferido, o promotor é um promotor de celobiohidrolase I (cbh1). Em um aspecto ainda mais preferido, o promotor é um promotor do gene celobiohidrolase I de *Trichoderma reesei* (cbh1). Em um aspecto o mais preferido, o promotor é o promotor de *Trichoderma reesei* CbM de nucleotídeos 505 a 1501 de SEQ ID NO 29 (GENBANK™ número de acesso D86235). Em um outro aspecto o mais preferido, o promotor é um promotor de celobiohidrolase II (cbh2). Em ainda outro aspecto mais preferido, o promotor é um promotor do gene de celobiohidrolase II de *Trichoderma reesei* (cbh2). Em outro aspecto o mais preferido, o promotor é o promotor de *Trichoderma reesei* cbh2 de nucleotídeos 1 a 582 de SEQ ID NO 30 (GENBANK™ número de acesso M55080).

Em outro aspecto preferido, o promotor é o promotor de NA2-tpi. Em outro aspecto preferido, o promotor é um promotor de TAKA amilase. Em outro aspecto preferido, o promotor é um promotor de amiloglicosidase de *Fusarium venenatum*. Em outro aspecto preferido, o promotor é um promotor de protease tipo tripsina de *Fusarium oxysporum*. Em outro aspecto preferido, o promotor é um promotor de glucoamilase (glaA) de *Aspergillus niger* ou *Aspergillus awamori*.

Em um aspecto preferido, a região de promotor dirige a expressão do primeiro, segundo e terceiro polinucleotídeos, e alternativamente também o quarto polinucleotídeo.

Terminadores

5 O terminador pode ser qualquer sequência de terminador de transcrição reconhecida por uma célula hospedeira para terminar a transcrição. A sequência do terminador é ligada operativamente ao término 3' da sequência de nucleotídeos codificando o polipeptídeo. Qualquer terminador que é funcional na célula hospedeira de escolha pode ser usado na
10 presente invenção.

Os terminadores preferidos para células hospedeiras de fungo filamentoso são obtidos a partir dos genes para TAKA amilase de *Aspergillus oryzae*, glucoamilase de *Aspergillus niger*, antranilato sintase de *Aspergillus nidulans*, alfa-glicosidase de *Aspergillus niger*, e protease tipo tripsina de
15 *Fusarium oxysporum*.

Os terminadores preferidos para células hospedeiras de levedura são obtidos a partir de genes para enolase de *Saccharomyces cerevisiae*, citocromo *Saccharomyces cerevisiae* C (CYC1), e gliceraldeído-3-fosfato desidrogenase de *Saccharomyces cerevisiae*. Outros terminadores
20 utilizáveis para células hospedeiras de levedura são descritos por Romanos et al, 1992, supra.

Em um aspecto preferido, o terminador é um terminador de gene de celobiohidrolase. Em um aspecto mais preferido, o terminador é um terminador de gene de celobiohidrolase I (*cbh1*). Em um aspecto ainda mais
25 preferido, o terminador é um terminador de gene de celobiohidrolase I de *Trichoderma reesei* (*cbh1*). Em um aspecto o mais preferido, o terminador é o terminador *cbh1* de *Trichoderma reesei* de SEQ ID NO 31. Em um outro aspecto o mais preferido, o terminador é um terminador de gene celobiohidrolase II(*cbh2*). Em ainda outro aspecto mais preferido, o

terminador é um terminador de gene de celobiohidrolase II de *Trichoderma reesei* (*cbh2*). Em outro aspecto o mais preferido, o terminador é o terminador *cbh2* de SEQ ID NO 32 de *Trichoderma reesei*.

5 Em outro aspecto preferido, o terminador é um terminador de TAKA amilase. Em outro aspecto preferido, o promotor é a terminador de protease de tipo tripsina de *Fusarium oxysporum*. Em outro aspecto preferido, o promotor é um terminador de glucoamilase (*glaA*) de *Aspergillus niger* ou *Aspergillus awamori*.

Outras sequências regulatórias

10 As construções de proteína de fusão podem ainda compreender outros elementos regulatórios, como um líder, sequência de poliadenilação e outros elementos.

A sequência regulatórios também pode ser uma sequência líder apropriada, uma região não traduzida de um mRNA que é importante para 15 tradução pela célula hospedeira. A sequência líder é ligada operativamente ao término 5' da sequência de nucleotídeos codificando uma proteína de fusão. Qualquer sequência líder que é funcional na célula hospedeira de escolha pode ser usada na presente invenção.

20 Os líderes preferidos para células hospedeiras de fungo filamentoso são obtidos a partir de genes para TAKA amilase de *Aspergillus oryzae* e triose fosfato isomerase de *Aspergillus nidulans*.

Os líderes apropriados para células hospedeiras de levedura são obtidos a partir de genes para enolase de *Saccharomyces cerevisiae* (ENO-1) 3- fosfoglicerato quinase de *Saccharomyces cerevisiae*, alfa-fator de 25 *Saccharomyces cerevisiae*, e álcool desidrogenase/gliceraldeído-3-fosfato desidrogenase de *Saccharomyces cerevisiae* e (ADH2/GAP).

A sequência regulatória também pode ser uma sequência de poliadenilação, uma sequência ligada operativamente ao término 3' da sequência de nucleotídeos e, quando transcrita, é reconhecida pela célula

hospedeira como um sinal para adicionar resíduos de poliadenosina a mRNA transcrito. Qualquer sequência de poliadenilação que é funcional na célula hospedeira de escolha pode ser usada na presente invenção.

5 As sequências de poliadenilação preferidas para células hospedeiras de fungo filamentoso são obtidas dos genes para TAKA amilase de *Aspergillus oryzae*, glucoamilase de *Aspergillus niger*, antranilato sintase de *Aspergillus nidulans*, protease tipo tripsina de *Fusarium oxysporum*, e alfa-glucosidase de *Aspergillus niger*.

10 As sequências de poliadenilação utilizáveis para células hospedeiras de levedura são descritas por Guo e Sherman, 1995, *Molecular Cellular Biology*, 15 5983-5990.

15 Pode ser desejável adicionar sequências regulatórias que permitem a regulação da expressão do polipeptídeo com relação ao crescimento da célula hospedeira. Os exemplos de sistemas regulatórios são os que causam a expressão do gene a ser ligado ou desligado em resposta a um estímulo químico ou físico, incluindo a presença de um composto regulatório. Os sistemas regulatórios em sistemas procarióticos incluem os sistemas de operador *lac*, *tac*, e *trp*. Em levedura, o sistema ADH2 ou sistema GAL1 pode ser usado. Em fungo filamentoso, o promotor de TAKA alfa-amilase, promotor glucoamilase de *Aspergillus niger*, e promotor glucoamilase de *Aspergillus oryzae* podem ser usados como sequências regulatórias. Outros exemplos de sequências regulatórias são os que permitem a amplificação do gene. Em sistemas eucarióticos, estes incluem o gene dihidrofolato reductase, que é amplificado na presença de r metotrexato e genes de metalotioneina que são amplificados em metais pesados. Nestes casos, a sequência de nucleotídeos codificando o polipeptídeo deve ser ligada operativamente com a sequência regulatória.

20
25

As construções de proteína de fusão contém preferivelmente um ou mais marcadores selecionáveis que permitem a seleção fácil de células

transformadas, transfectadas, transduzidas ou outras. Um marcador selecionável é um gene cujo produto provê resistência a biocida ou viral, resistência a metais pesados, prototrofia a auxotróficos, e outros.

Exemplos de marcadores bacterianos selecionáveis são os genes *dal* de *Bacillus subtilis* ou *Bacillus licheniformis*, ou marcadores que conferem resistência a antibióticos como ampicilina, canamicina, cloranfenicol, ou resistência a tetraciclina. Os marcadores apropriados para células hospedeiras de levedura são ADE2, HIS3, LEU2, LYS2, MET3, TRP1, e URA3. Os marcadores selecionáveis para uso em uma célula hospedeira de fungo filamentoso incluem, mas não são limitados a, *amdS* (acetamidase), *argB* (ornitina carbamoiltransferase), *bar* (fosfinotricina acetiltransferase), *hph* (higromicina fosfotransferase), *niaD* (nitrato reductase), *pyrG* (orotidina-5'- fosfato decarboxilase), *sC* (sulfato adeniltransferase), e *trpC* (antranilato sintase), assim como equivalentes dos mesmos. Preferidos para uso em uma célula de *Aspergillus* são os genes *amdS* e *pyrG* de *Aspergillus nidulans* ou *Aspergillus oryzae* e o gene *bar* de *Streptomyces hygroscopicus*.

Vetores de expressão

A presente invenção também refere-se aos vetores de expressão recombinante compreendendo um polinucleotídeo codificando uma proteína de fusão, um promotor e sinais de parada transcripcional e translacional. Os vários ácidos nucleicos e sequências de controle descritos aqui podem ser unidos juntos para produzir um vetor de expressão recombinante que pode incluir um ou mais sítios de restrição convenientes para permitir a inserção ou substituição da sequência de nucleotídeo codificando o polipeptídeo em tais sítios. Alternativamente, um polinucleotídeo codificando uma proteína de fusão pode ser expressado por inserção da sequência de nucleotídeo ou uma construção de proteína de fusão compreendendo a sequência em um vetor apropriado para expressão. Ao criar

o vetor de expressão, a sequência de codificação está localizada no vetor de modo que a sequência de codificação é ligada operativamente com as sequências de controle apropriadas para expressão.

5 O vetor de expressão recombinante pode ser qualquer vetor (por exemplo, um plasmídeo ou vírus) que pode ser convenientemente submetido a procedimentos de DNA recombinantes e pode ocasionar a expressão da sequência de nucleotídeo. A escolha do vetor tipicamente irá depender da compatibilidade do vetor com a célula hospedeira em que o vetor deve ser introduzido. Os vetores podem ser lineares ou plasmídeos circulares
10 fechados.

O vetor pode ser um vetor autonomamente replicante, isto é, um vetor que existe como uma entidade extracromossômica, cuja replicação é independente da replicação cromossômica, por exemplo, um plasmídeo, um elemento extracromossômico, um minicromossomo, ou um cromossomo artificial. O vetor pode conter qualquer meio para assegurar uma auto-replicação. Alternativamente, o vetor pode ser um que, quando introduzido na
15 célula hospedeira, é integrado no genoma e replicado junto com o (s) cromossomo (s) em que ele foi integrado. Além disso, um vetor único ou plasmídeo ou dois ou mais vetores ou plasmídeos que juntos contém o DNA
20 total a ser introduzido no genoma da célula hospedeira, ou um transposon pode ser usado.

Os vetores da presente invenção preferivelmente contêm um ou mais marcadores selecionáveis que permitem uma seleção fácil de células transformadas, transfectadas, transduzidas, ou semelhantes. Exemplos de
25 marcadores selecionáveis de fungo filamentosos, leveduras, bacterianos são descritos aqui.

Um vetor da presente invenção preferivelmente contém um elemento(s) que permite a integração do vetor no genoma da célula hospedeira ou replicação autônoma do vetor na célula independente do

genoma.

Para integração no genoma de célula hospedeira, o vetor pode se basear na sequência de polinucleotídeo codificando o polipeptídeo ou qualquer outro elemento do vetor para integração no genoma por recombinação homóloga ou não homóloga. Alternativamente, o vetor pode conter sequências de nucleotídeos adicionais para dirigir a integração por recombinação homóloga no genoma da célula hospedeira em um local (s) preciso no (s) cromossomo (s). Para aumentar a probabilidade de integração em um local preciso, os elementos integracionais devem conter preferivelmente um número suficiente de ácidos nucleicos, como 100 a 10.000 pares de base, preferivelmente 400 a 10.000 pares de base, e o mais preferivelmente 800 a 10.000 pares de base, tendo um grau elevado de identidade com a sequência de marcação correspondente para melhorar a probabilidade de recombinação homóloga. Os elementos integracionais podem ser qualquer sequência, isto é, homóloga com a sequência de marcação no genoma da célula hospedeira. Além disso, os elementos integracionais podem ser as sequências de nucleotídeo de não codificação e de codificação. Por outro lado, o vetor pode ser integrado no genoma da célula hospedeira por recombinação não homóloga.

Para replicação autônoma, o vetor pode ainda compreender uma origem de replicação permitindo ao vetor replicar autonomamente na célula hospedeira em questão. A origem de replicação pode ser qualquer replicador de plasmídeo mediando a replicação autônoma que funciona em uma célula. O termo “origem de replicação ” ou “replicador de plasmídeo” é definido aqui como uma sequência de nucleotídeo que permite a um plasmídeo ou vetor replicar *in vivo*.

Exemplos de origens bacterianas de replicação são as de origens de replicação de plasmídeos pBR322, pUC19, pACYC177, e pACYC184 permitindo a replicação em *E. coli*, e pUB110, pE194, pTA1060,

e pAM β 1 permitindo a replicação em *Bacillus*.

Exemplos de origens de replicação para uso em uma célula hospedeira de levedura são a origem de replicação de 2 microns, ARS1, ARS4, a combinação de ARS1 e CEN3, e a combinação de ARS4 e CEN6.

5 Exemplos de origens de replicação utilizáveis em uma célula de fungo filamentosos são AMA1 e ANS1 (Gems *et al.*, 1991, *Gene* 98: 61 - 67; Cullen *et al.*, 1987, *Nucleic Acids Research* 15: 9163 - 9175; WO 00/24883). Isolamento do gene AMA1 e construção de plasmídeo ou vetores compreendendo o gene podem ser alcançados de acordo com os métodos
10 descritos em WO 00/24883.

Mais do que uma cópia de um polinucleotídeo codificando uma proteína de fusão pode ser inserida em uma célula hospedeira para aumentar a produção do produto de gene. Um aumento no número de cópia do polinucleotídeo pode ser obtido por integração de pelo menos uma cópia
15 adicional da sequência no genoma de célula hospedeira ou por inclusão de um gene de marcador selecionável amplificável com o polinucleotídeo onde células contendo cópias amplificadas do gene de marcador selecionável, e assim cópias adicionais do polinucleotídeo, podem ser selecionadas por cultivo das células na presença do agente selecionável apropriado.

20 Os procedimentos usados para ligar os elementos descritos acima para construir os vetores de expressão recombinante da presente invenção são bem conhecidos do versado na técnica (ver, por exemplo, Sambrook *et al.*, 1989, *supra*).

Células hospedeiras

25 A presente invenção também refere-se a células hospedeiras de fungos recombinantes, compreendendo um polinucleotídeo codificando uma proteína de fusão da presente invenção, que são usados com vantagem na produção recombinante da proteína. Um vetor compreendendo um polinucleotídeo da presente invenção é introduzido em uma célula hospedeira

fúngica de modo que o vetor é mantido como um integrante cromossômico ou como um vetor extra-cromossômico auto-replicante como descrito acima. O termo "célula hospedeira" engloba qualquer progênie ou célula parental que não é idêntica à célula parental devido às mutações que ocorrem durante a
5 replicação.

A célula hospedeira fúngica pode ser qualquer célula fúngica utilizável na produção recombinante de um polipeptídeo da presente invenção.

"Fungos", como usado aqui, incluem os filós Ascomycota, Basidiomycota, Chytridiomycota, e Zygomycota (como definidos por
10 Hawksworth *et al.*, In, *Ainsworth and Bisby's Dictionary of The Fungi*, 8^a. edição, 1995, CAB International, University Press, Cambridge, UK) assim como Oomycota (como citado em Hawksworth *et al.*, 1995, *supra*, página 171) e todos os fungos mitospóricos (Hawksworth *et al.*, 1995, *supra*).

Em um aspecto mais preferido, a célula hospedeira de fungo é uma célula de levedura. "Levedura" como usada aqui inclui levedura ascosporogenosa (Endomicetales), levedura basidiosporogenosa, e levedura pertencendo aos Fungi imperfecti (Blastomicetes). Porque a classificação de levedura pode mudar no futuro, para os fins desta invenção, levedura deve ser
15 definida como descrito em *Biology and Activities of Yeast* (Skinner, F. A., Passmore, S. M., e Davenport, R. R., eds, *Soc. App. Bacteriol. Symposium* série n^o. 9, 1980).

Em um aspecto ainda mais preferido, a célula hospedeira de levedura é uma célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*,
25 *Saccharomyces*, *Schizosaccharomyces*, ou *Yarrowia*.

Em um aspecto o mais preferido, a célula hospedeira de levedura é uma célula de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, ou *Saccharomyces*

oviformis. Em outro aspecto o mais preferido, a célula hospedeira de levedura é uma célula de *Kluyveromyces lactis*. Em outro aspecto o mais preferido, a célula hospedeira de levedura é uma célula de *Yarrowia lipolytica*.

Em outro aspecto mais preferido, a célula hospedeira de fungo é uma célula de fungo filamentoso. “Fungos filamentosos” incluem todas as formas filamentosas da subdivisão Eumycota e Oomycota (como definido por Hawksworth *et al.*, 1995, *supra*). Os fungos filamentosos geralmente são caracterizados por uma parede miceliana composta de quitina, celulose, glucano, quitosana, manano, e outros polissacarídeos complexos. Crescimento vegetativo é por alongamento das hifas e catabolismo de carbono é obrigatoriamente aeróbico. Em contraste, crescimento vegetativo por leveduras como *Saccharomyces cerevisiae* é por formação de broto de um talo unicelular e catabolismo de carbono pode ser fermentativo.

Em um aspecto mais preferido, a célula hospedeira de fungo filamentoso é uma célula de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Chrysosporium*, *Coprinus*, *Conolus*, *Cryptococcus*, *Fihbasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paealomyces*, *Penicillium*, *Phanerochaete*, *Phtebia*, *Piromyoes*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolyptocladium*, *Trametes*, ou *trichoderma*. Em um aspecto o mais preferido, célula hospedeira de fungo filamentoso é uma célula de *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* ou *Aspergillus oryzae*. Em outro aspecto o mais preferido, célula hospedeira de fungo filamentoso é uma célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium grammearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambuanum*, *Fusarium sarochroum*, *Fusarium sporotnchioides*, *Fusarium sulphureum*,

Fusarium torulosum, *Fusarium tnchotheaoides*, ou *Fusarium venenatum*. Em outro aspecto o mais preferido, a célula hospedeira de fungo filamentoso é uma célula de *Bjerkandera adusta*, *Cenporiopsis aneirina*, *Cenporiopsis aneirina*, *Cenponopsis caregiea*, *Cenporiopsis gilvescens*, *Cenporiopsis*
 5 *pannoanta*, *Cenponopsis nvulosa* *Cenporiopsis subrufa*, *Cenponopsis subvermispora*, *Chrysosporium keratophilum*, *Chrysosporium lucknowense*, *Chrysosporium troucum*, *Chrysosponum merdanum*, *Chrysosporium mops*, *Chrysosponum pannicola*, *Chrysosporium queenslandicum*, *Chrysosponum zonatum*, *Coprinus cinereus*, *Conolus hirsutus*, *Humicola insolens*, *Humicola*
 10 *lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Pemallium purpurogenum*, *Phanerochaete chrysosponum*, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichodenva harzianum*, *Tnchodenna konmgii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, ou *Trichodenva vinde*.

15 As células fúngicas podem ser transformadas por um processo envolvendo a formação de protoplastos, transformação de protoplastos, e regeneração da parece celular em um modo conhecido. Os procedimentos apropriados para a transformação de células hospedeiras de *Aspergillus* e *Trichoderma* são descritos em EP 238 023 e Yelton *et al.*, 1984, *Proceedings*
 20 *of the National Academy of Sciences USA* 81: 1470 - 1474. Os métodos apropriados para transformar a espécie *Fusarium* são descritos por Malardier *et al.*, 1989, *Gene* 78: 147 - 158, e WO 96/00787. A levedura pode ser transformada usando os procedimentos descritos por Becker e Guarente, In
 25 *Abelson, J. N. e Simon, M. I., editores, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, volume 194, pp 182 - 187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; e Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920.

Produção e recuperação

Nos métodos da presente invenção, a célula hospedeira fúngica é cultivada em um meio nutriente apropriado para a produção de um polipeptídeo tendo atividade biológica usando métodos bem conhecidos na arte. Por exemplo, a célula pode ser cultivada por cultivo em frasco de
5 agitação, e fermentação em grande escala ou pequena escala (incluindo fermentações em estado contínuo, batelada, batelada alimentada, ou sólido) em laboratório ou fermentadores industriais realizada em um meio apropriado e sob condições permitindo ao polipeptídeo ser expresso e/ou isolado. O cultivo ocorre em um meio nutriente apropriado compreendendo fontes de
10 carbono e nitrogênio e sais inorgânicos, usando procedimentos conhecidos na arte. Os meios apropriados são disponíveis de fornecedores comerciais ou podem ser preparados de acordo com as composições publicadas (por exemplo, em catálogos do American Type Culture Collection).

Nos métodos da presente invenção, o polipeptídeo tendo
15 atividade biológica é selecionado dentre um grupo incluindo uma proteína de fusão, componentes da proteína de fusão e uma combinação da proteína de fusão e componentes da mesma.

Em um aspecto preferido, o polipeptídeo tendo atividade biológica é a proteína de fusão.

20 Em outro aspecto preferido, o polipeptídeo tendo atividade biológica é um componente(s) de uma proteína de fusão.

Em outro aspecto preferido, o polipeptídeo tendo atividade biológica é uma combinação de uma proteína de fusão e componentes da mesma.

25 Os polipeptídeos tendo atividade biológica podem detectados usando métodos conhecidos na técnica que são específicos para os polipeptídeos. Estes métodos de detecção podem incluir uso de anticorpos específicos, formação de um produto de enzima, ou desaparecimento de um substrato de enzima. Por exemplo, um teste de enzima pode ser usado para

determinar a atividade do polipeptídeo como descrito aqui, que pode incluir atividade de endoglucanase e uma atividade biológica específica.

O polipeptídeo resultante tendo atividade biológica, por exemplo proteína de fusão beta-glucosidase ou um componente da mesma, pode ser recuperado usando métodos conhecidos na arte. Por exemplo, o polipeptídeo pode ser recuperado do meio nutriente por procedimentos convencionais incluindo, mas não limitados a, centrifugação, filtração, extração, secagem por pulverização, evaporação, ou precipitação.

Os polipeptídeos da presente invenção podem ser purificados por uma variedade de procedimentos conhecidos na técnica incluindo, mas não limitados a, cromatografia (por exemplo, troca iônica, afinidade, hidrofóbico, cromatofocalização, e exclusão por tamanho), procedimentos eletroforéticos (por exemplo, focalização isoeétrica preparativa), solubilidade diferencial (por exemplo, precipitação por sulfato de amônio), SDS-PAGE, ou extração (ver, por exemplo, *Protein Purification*, J. -C, Janson e Lars Ryden, editores, VCH Publishers, New York, 1989) para obter polipeptídeos substancialmente puros.

Composições

A presente invenção também refere-se às composições compreendendo um polipeptídeo tendo atividade biológica da presente invenção. Preferivelmente, as composições são enriquecidas em tal polipeptídeo. O termo "enriquecido" indica que a atividade de endoglucanase da composição foi aumentada, por exemplo, com um fator de enriquecimento de pelo menos 1.1.

A composição pode compreender um polipeptídeo da presente invenção como o componente enzimático principal, por exemplo, uma composição mono-componente. Alternativamente, a composição pode compreender atividades enzimáticas múltiplas, como um aminopeptidase, amilase, carbohidrase, carboxipeptidase, catalase, celulase, quitinase,

cutinase, ciclodextrina glicosiltransferase, deoxiribonuclease, esterase, alfa-galactosidase, beta-galactosidase, glucoamilase, alfa-glucosidase, beta-glucosidase, haloperoxidase, invertase, lacase, lipase, manosidase, oxidase, enzima pectinolítica, peptidoglutaminase, peroxidase, fitase, polifenoloxidase, 5 enzima proteolítica, ribonuclease, transglutaminase, ou xilanoase. A(s) enzima (s) adicional(ais) pode(m) ser produzida(s), por exemplo, por um microorganismo pertencendo no gênero *Aspergillus*, preferivelmente *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, ou *Aspergillus oryzae*; *Fusarium*, preferivelmente *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sulphureum*, *Fusarium toruloseum*, *Fusarium trichothecioides*, ou *Fusarium venenatum*; *Humicola*, preferivelmente *Humicola insolens* ou *Humicola lanuginosa*; ou *Trichoderma*, preferivelmente *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, ou *Trichoderma viride*.

20 As composições de polipeptídeo podem ser preparadas de acordo com os métodos conhecidos na técnica e podem estar na forma de um líquido ou uma composição em pó. Por exemplo, a composição de polipeptídeo pode estar na forma de um granulado ou um microgranulado. O polipeptídeo a ser incluído na composição pode ser estabilizado de acordo 25 com os métodos conhecidos na arte.

Exemplos são dados abaixo de usos preferidos das composições de polipeptídeo da invenção. A dosagem da composição de polipeptídeo da invenção e outras condições sob as quais a composição é usada podem ser determinadas com base em métodos conhecidos na arte.

Usos

A presente invenção é também dirigida a métodos para usar as proteínas de fusão ou componentes da mesma ou composições da mesma.

Métodos de processamento de material celulósico

5 Os métodos da presente invenção são particularmente utilizáveis para melhorar a secreção de polipeptídeos tendo atividade celulolítica ou hemicelulolítica em quantidades comercialmente importantes, que podem ser usadas para degradar ou converter material lignocelulósico. Estes polipeptídeos incluem, mas não são limitados a endoglucanases, 10 celobiohidrolases, beta-glucosidases, xilanases, beta-xilosidases, arabinofuranosidases, acetil xilano I esterases, e esterase de ácido ferúlicos. Consequentemente, a presente invenção também refere-se a métodos para degradar ou converter um material celulósico, compreendendo: tratar o material celulósico com uma quantidade eficaz de uma composição de enzima 15 celulolítica na presença de uma quantidade eficaz de uma proteína de fusão ou um componente da mesma tendo atividade celulolítica ou hemicelulolítica obtida de acordo com os presentes métodos. Para fins de ilustração, um polipeptídeo tendo atividade beta- glucosidase obtido de acordo com os métodos da presente invenção, por exemplo uma proteína de fusão de beta- 20 glucosidase ou um componente da mesma, é usado para fins ilustrativos.

Biomassa celulósica pode incluir, mas não é limitada a recursos de madeira, refugo sólido municipal, papel usado, colheitas, e resíduos de colheitas (ver, por exemplo, Wiselogel *et al.*, 1995, in Handbook on Bioethanol (Charles E. Wyman, editor), pp. 105 - 118, Taylor & Francis, 25 Washington D. C.; Wyman, 1994, *Bioresource Technology* 50: 3 - 16; Lynd, 1990, *Applied Biochemistry and Biotechnology* 24/25: 695 - 719; Mosier *et al.*, 1999, Recent Progress in Bioconversion of Lignocellulosics, in *Advances in Biochemical Engineering/Biotechnology*, T. Scheper, editor responsável, volume 65, pp. 23 - 40, Springer-Verlag, New York).

O polissacarídeo predominante na parede celular primária de biomassa é celulose, o segundo o mais abundante é hemi-celulose, e o terceiro é pectina. A parede celular secundária, produzida depois da célula ter parado de crescer, também contém polissacarídeos e é reforçada por lignina polimérica covalentemente reticulada em hemicelulose. Celulose é um homopolímero de anidrocetobiose e assim um beta-(1-4)-D-glucano linear, enquanto hemiceluloses incluem uma variedade de compostos, como xilanos, xiloglucanos, arabinoxilanos, e mananos em estruturas ramificadas complexas com um espectro de substituintes. Apesar de geralmente polimorfa, celulose é encontrada em tecido de planta primariamente como uma matriz cristalina insolúvel de cadeias glucano paralelas. Hemiceluloses geralmente ligam por ponte hidrogênio a celulose, assim como a outras hemiceluloses, que ajudam a estabilizar a matriz da parede celular. Será entendido aqui que o termo "material celulósico" ou "celulose" também engloba lignocelulose.

Nos métodos da presente invenção, a composição de enzima celulolítica pode compreender qualquer proteína envolvida no processamento de material celulósico em glicose, ou hemicelulose em xilose, mannose, galactose, e arabinose, seus polímeros, ou produtos derivados dos mesmos como descrito abaixo. A composição de enzima celulolítica pode ser uma preparação de mono-componente, por exemplo uma endoglucanase, uma preparação de múltiplos componentes, por exemplo, endoglucanase, celobiohidrolase, beta-glucosidase, ou uma combinação de preparações de proteína de multicomponentes e monocomponente. As proteínas celulolíticas pode ter atividade, isto é hidrolisar celulose, ou na faixa de pH ácido, neutro ou alcalino. A composição de enzima celulolítica pode ainda compreender um polipeptídeo tendo atividade melhoradora celulolítica de acordo com WO 2005/074647, WO 2005/074656, e pedido publicado US 2007/0077630

A proteína celulolítica pode ser de origem de fungo ou bacteriana, que pode ser obtível ou isolada e purificada de microorganismos

que são conhecidos como capazes de produzir enzimas celulolíticas, por exemplo espécies de *Bacillus*, *Pseudomonas*, *Humicola*, *Coprinus*, *Thielavia*, *Fusarium*, *Myceliophthora*, *Acremonium*, *Cephalosporium*, *Scytalidium*, *Penicillium* ou *Aspergillus* (ver, por exemplo, EP 458162), especialmente as

5 produzidas por uma cepa selecionada dentre *Humicola insolens* (reclassificada como *Scytalidium thermophilum*, ver por exemplo, Patente US No 4.435.307), *Coprinus cinereus*, *Fusarium oxysporum*, *Myceliophthora theimophila*, *Meripilus giganteus*, *Thielavia terrestris*, *Acremonium sp*, *Acremonium persicinum*, *Acremonium acremonium*, *Acremonium*

10 *brachyphenium*, *Acremonium dichromosporum*, *Acremonium obclavatum*, *Acremonium pinkertoniae*, *Acremonium roseogriseum*, *Acremonium incoloratum*, e *Acremonium furatum*, preferivelmente de *Humicola insolens* DSM 1800, *Fusarium oxysporum* DSM 2672, *Myceliophthora thermophila* CBS 117 65, *Cephalosporium sp* RYM-202, *Acremonium sp* CBS 47894,

15 *Acremonium sp* CBS 265 95, *Acremonium persicinum* CBS 169 65, *Acremonium acremonium* AHU 9519, *Cephalosporium sp* CBS 535 71, *Acremonium brachyphenium* CBS 866 73, *Acremonium dichromosporium* CBS 683 73, *Acremonium obclavatum* CBS 311 74, *Acremonium pmkertoniae* CBS 157 70, *Acremonium roseogriseum* CBS 134 56, *Acremonium*

20 *incoloratum* CBS 146 62, e *Acremonium furatum* CBS 29970H. As proteínas celulolíticas também podem ser obtidas a partir de *Trichoderma* (particularmente *Trichoderma vinde*, *Trichoderma reesei*, e *Trichoderma koningii*), *Bacillus* alcalofílico (ver, por exemplo, Patente US No 3.844.890 e EP 458162), e *Streptomyces* (ver, por exemplo, EP 458162). Os mutantes

25 quimicamente modificados ou engenheirados em proteínas de proteínas celulolíticas também podem ser usados.

As proteínas celulolíticas especialmente apropriadas são as celulasas descritas em EP 495.257, EP 531.372, WO 96/11262, WO 96/29397, WO 98/08940. Outros exemplos são variantes de celulase como as

descritas em WO 94/07998, EP 531.315, Patente US No 4.435.307, Patente US No 5.457.046, Patente US No 5.648.263, Patente US No 5.686.593, Patente US No 5.691.178, Patente US No 5.763.254, Patente US No 5.776.757, WO 89/09259, WO 95/24471, WO 98/12307, e PCT/DK98/00299.

5 Como mencionado acima, as proteínas celulolíticas usadas nos métodos da presente invenção podem ser preparações de monocomponente, isto é, um componente essencialmente isento de outro componentes celulolíticos. O componente único pode ser um componente recombinante, isto é, produzindo por clonagem de uma sequência de DNA codificando o
10 componente único e subsequentemente a célula transformada com a sequência de DNA e expressada no hospedeiro (ver, por exemplo, WO 91/17243 e WO 91/17244). O hospedeiro é preferivelmente um hospedeiro heterólogo (enzima é estranha para o hospedeiro), mas o hospedeiro pode sob algumas condições também ser um hospedeiro homólogo (enzima é nativa para o
15 hospedeiro). As proteínas celulolíticas de monocomponente também podem ser preparadas por purificação desta proteína a partir do caldo de fermentação.

 As proteínas celulolíticas usadas nos métodos da presente invenção podem ser produzidas por fermentação das cepas microbianas acima notadas em um meio nutriente contendo fontes de carbono e nitrogênio
20 apropriadas e sais inorgânicos, usando procedimentos conhecidos na arte (ver, por exemplo, Bennett, J W e LaSure, L (eds), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991). Os meios apropriados são disponíveis de fornecedores comerciais ou podem ser preparados de acordo com composições publicadas (por exemplo, nos catálogos do American Type
25 Culture Collection). As faixas de temperatura e outras condições apropriadas para crescimento e produção de proteínas celulolíticas são bem conhecidas na arte (ver, por exemplo, Bailey, J E, e Ollis, D F, *Biochemical Engineering Fundamentals*, McGraw-Hill Book Company, NY, 1986).

 A fermentação pode ser qualquer método de cultivo de uma

célula resultante na expressão ou isolamento de uma proteína celulolítica. A fermentação porem, assim, ser entendida como compreendendo cultivo em frasco de agitação, e fermentação em grande escala ou pequena escala (incluindo fermentações em contínuo, batelada, batelada alimentada, ou estado sólido) em laboratório ou fermentadores industriais realizada em um meio apropriado e sob condições permitindo à proteína celulolítica ser expressada e/ou isolada. As proteínas celulolíticas resultantes produzidas pelos métodos descritos acima podem ser recuperadas do meio de fermentação e purificadas por procedimentos convencionais como aqui descrito.

Os exemplos de preparações de enzima celulolítica comerciais apropriados para uso na presente invenção incluem, por exemplo,, CELLUCLAST™ (disponível de Novozymes A/S) e NOVOZYM™ 188 (disponível de Novozymes A/S). Outras preparações comercialmente disponíveis compreendendo celulase que podem ser usadas incluem CELLUZYME™, CEREFLO™ e ULTRAFLO™ (Novozymes A/S), LAMINEX™ e SPEZYME™ CP (Genencor Int.), ROHAMENT™ 7069 W (Rohm GmbH), e FIBREZYME® LDI, FIBREZYME® LBR, ou VISCOSTAR® 150L (Dyadic International, Inc, Jupiter, FL, USA) As enzimas celulase são adicionadas em quantidades eficazes de cerca de 0,001% a cerca de 5,0 % em peso de sólidos, mais preferivelmente de cerca de 0,025% a cerca de 4,0% em peso de sólidos, e o mais preferivelmente de cerca de 0,005% a cerca de 2,0% em peso de sólidos.

As proteínas celulolíticas resultantes ou proteínas de beta-glucosidase produzidas pelos métodos acima mencionados podem ser recuperadas do meio de fermentação por procedimentos convencionais incluindo, mas não limitados a centrifugação, filtração, secagem por pulverização, evaporação ou precipitação. A proteína recuperada pode ser ainda purificada por vários procedimentos de cromatografia, por exemplo

cromatografia de troca iônica, cromatografia de filtração de gel, cromatografia de afinidade, ou outros.

A atividade da enzima celulolítica pode ser determinada usando qualquer método bem conhecido na arte.

5 Os exemplos de preparações celulolíticas apropriados para uso na presente invenção incluem, por exemplo, CELLUCLAST™ (disponível de Novozymes A/S) e NOVOZYM™ 188 (disponível de Novozymes A/S). Outras preparações comercialmente disponíveis compreendendo celulase que podem ser usadas incluem CELLUZYME™, CEREFLO™ e ULTRAFLO™
10 (Novozymes A/S), LAMINEX™ e SPEZYME™ CP (Genencor Int), e ROHAMENT™ 7069 W (Rohm GmbH). As enzimas celulase são adicionadas em quantidades eficazes de cerca de 0,001% a cerca de 5,0 % em peso de sólidos, mais preferivelmente de cerca de 0,025% a cerca de 4,0% em peso de sólidos, e o mais preferivelmente de cerca de 0,005% a cerca de 2,
15 0% em peso de sólidos.

Como acima mencionado, as proteínas celulolíticas usadas nos métodos da presente invenção podem ser preparações de monocomponente, isto é, um componente essencialmente isento de outros componentes celulolíticos. O componente único pode ser um componente recombinante,
20 isto é, produzindo por clonagem de uma sequência de DNA codificando o componente único e subsequentemente a célula transformada com a sequência de DNA e expressada no hospedeiro (ver, por exemplo, WO 91/17243 e WO 91/17244). Outros exemplos proteínas celulolíticas mono componente incluem, mas não limitados aos descritos em JP-07203960-A e WO-9206209.
25 O hospedeiro é preferivelmente um hospedeiro heterólogo (enzima é estranha para o hospedeiro), mas o hospedeiro pode sob algumas condições também ser um hospedeiro homólogo (enzima é nativa para o hospedeiro). As proteínas celulolíticas de monocomponente também podem ser preparadas por purificação desta proteína a partir do caldo de fermentação.

Os exemplos de proteínas celulolíticas de monocomponente utilizáveis na prática dos métodos da presente invenção incluem, mas não são limitados a, endoglucanase, celobiohidrolase, e outras enzimas utilizáveis na degradação de biomassa celulósica.

5 O termo "endoglucanase" já foi definido aqui. O termo "celobiohidrolase" é definido aqui como uma 1,4-beta-D-glucano celobiohidrolase (E.C. 32 1 91), que catalisa a hidrólise de 1,4-beta-D-ligações glucosídicas, celooligosacarídeos, ou qualquer polímero contendo beta-1,4 ligada glicose, liberando celobiose a partir das extremidades
10 reductoras ou não reductoras da cadeia. Para fins da presente invenção, a atividade de celobiohidrolase pode ser determinada de acordo com os procedimentos descrito por Lever et al, 1972, *Anal Biochem* 47 273-279 e por van Tilbeurgh et al, 1982, *FEBS Letters* 149 152-156, van Tilbeurgh e Claeysens, 1985 *FEBS Letters* 187 283-288. Na presente invenção, o método
15 de Lever et al foi empregado para avaliar a hidrólise de celulose em forragem de milho.

Os polipeptídeos da presente invenção são usado em conjunto com proteínas celulolíticas para degradar o componente celulósico do substrato da biomassa, (ver, por exemplo, Brigham et al, 1995, em Handbook
20 on Bioethanol (Charles E Wyman, editor), pp 119-141, Taylor & Francis, Washington D C, Lee, 1997, *Journal of Biotechnology* 56 1-24).

As quantidades ótimas de um polipeptídeo tendo atividade de beta- glucosidase e outras proteínas celulolíticas dependem de vários fatores incluindo, mas não limitados a mistura de proteínas de componentes
25 celulolíticos, o substrato celulósico, a concentração de substrato celulósico, o pré-tratamento (s) do substrato celulósico, temperatura, tempo, pH, e inclusão de organismo de fermentação (por exemplo, levedura para sacarificação e fermentação simultâneas). O termo "proteína celulolíticas" é definido aqui como as proteínas ou misturas de proteínas mostradas como sendo capazes de

hidrolisar ou converter ou degradar celulose sob as condições testadas. Suas quantidades são geralmente medidas por um teste comum como BCA (ácido bicinconínico, P. K. Smith et al, 1985, *Anal Biochem* 150 76), e a quantidade preferida adicionada em proporção à quantidade de biomassa sendo

5 hidrolisada.

Em um aspecto preferido, a quantidade de polipeptídeo tendo atividade beta-glucosidase por g de material celulósico é cerca de 0,01 a cerca de 2,0 mg, preferivelmente cerca de 0,025 a cerca de 1,5 mg, mais preferivelmente cerca de 0,05 a cerca de 1,25 mg, mais preferivelmente cerca

10 de 0,075 a cerca de 1,25 mg, mais preferivelmente cerca de 0,1 a cerca de 1,25 mg, ainda mais preferivelmente cerca de 0,15 a cerca de 1,25 mg, e o mais preferivelmente cerca de 0,25 a cerca de 1,0 mg por g de material celulósico.

Em outro aspecto preferido, a quantidade de proteínas celulolíticas por g de material celulósico é cerca de 0,5 a cerca de 50 mg, preferivelmente cerca de 0,5 a cerca de 40 mg, mais preferivelmente cerca de 0,5 a cerca de 25 mg, mais preferivelmente cerca de 0,75 a cerca de 20 mg, mais preferivelmente cerca de 0,75 a cerca de 15 mg, ainda mais preferivelmente cerca de 0,5 a cerca de 10 mg, e o mais preferivelmente cerca

15 de 2,5 a cerca de 10 mg por g de material celulósico.

20

Os métodos da presente invenção podem ser usados para degradar ou converter um material celulósico, por exemplo lignocelulose, em muitas substâncias utilizáveis, por exemplo, produtos químicos e combustíveis. Além disso, para etanol, algumas mercadorias e especialmente

25 produtos químicos especiais que podem ser produzidos a partir de celulose incluem xilose, acetona, acetato, glicina, lisina, ácidos orgânicos (por exemplo, ácido láctico), 1,3-propanodiol, butanodiol, glicerol, etileno glicol, furfural, polihidroxicanoatos, e ácido cis,cis-mucônico, (Lynd, L. R., Wyman, C. E., e Gerngross, T. U., 1999, *Biocommodity Engineering*,

Biotechnol. Prog., 15: 777 - 793; Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179 - 212, e Ryu, D. D. Y., e Mandels, M., 1980, Cellulases: biosynthesis and applications, *Enz. Microb. Technol.*, 2: 91 - 102). Benefícios de coprodução potenciais se estendem além da síntese de produtos orgânicos múltiplos de carboidrato fermentável. Resíduos ricos em lignina restantes após processamento biológico podem ser convertidos em produtos químicos derivados de lignina, ou usados para a produção de energia.

10 Métodos convencionais usados para processar o material celulósico de acordo com os métodos da presente invenção são bem entendidos pelo versado na técnica. Os métodos da presente invenção podem ser implementados usando qualquer aparelho de processamento de biomassa convencional configurado para operar de acordo com a invenção.

15 Este aparelho pode incluir um reator agitado de batelada, um reator agitado de fluxo contínuo com ultrafiltração, um reator de coluna de fluxo tampão contínuo (Fernanda de Castilhos Corazza, Flavio Faria de Moraes, Gisella Maria Zanin e Ivo Neitzel, 2003, Optimal control in fed-batch reactor for the celobiose hidrolisis, *Acta Scientiarum Technology* 25(1) 33-
20 38, Gusakov, A. V., e Sinitsyn, A. P., 1985, Kinetics of the enzymatic hydrolysis of cellulose: 1. A mathematical model for a batch reactor process, *Enz. Microb. Technol.* 7: 346 - 352), um reator por atrito (Ryu, S. K., e Lee, J. M., 1983, Biconversion of waste cellulose by using an attrition bioreactor, *Biotechnol. Bioeng.* 25: 53 - 65), ou um reator com agitação intensiva
25 induzida por um campo eletromagnético (Gusakov, A. V., Sinitsyn, A. P., Davydkin, I. Y., Davydkin, V. Y., Protas, O. V., 1996, Enhancement of enzymatic cellulose hydrolysis using a novel type of bioreactor with intensive stirring induced by electromagnetic field, *Appl. Biochem. Biotechnol.* 56: 141 - 153).

Os métodos convencionais incluem, mas não são limitados a, sacarificação, fermentação, hidrólise e fermentação separadas (SHF), sacarificação e fermentação simultâneas (SSF), sacarificação e cofermentação simultâneas (SSCF), hidrólise e fermentação híbridas (HHF), e conversão microbiana direta (DMC).

SHF usa etapas de processo separadas para primeiro hidrolisar enzimaticamente a celulose em glicose e então, em uma etapa subsequente, fermentar a glicose em etanol. Em SSF, a hidrólise enzimática de celulose e a fermentação de glicose em etanol é combinada em uma etapa (Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179 - 212). SSCF inclui a cofermentação de açúcares múltiplos (Sheehan, J., e Himmel, M., 1999, Enzymes, energy and the environment: A strategic perspective on the U.S. Department of Energy's research and development activities for bioethanol, *Biotechnol. Prog.* 15: 817 - 827). HHF inclui duas etapas separadas realizadas no mesmo reator mas em temperaturas diferentes, isto é, sacarificação enzimática de temperatura elevada seguido por SSF em uma temperatura inferior que a cepa de fermentação pode tolerar. DMC combina todos os três processos (produção de celulase, hidrólise de celulose, e fermentação) em uma etapa (Lynd, L. R., Weimer, P. J., van Zyl, W. H., e Pretorius, I. S., 2002, Microbial cellulose utilization: Fundamentals and biotechnology, *Microbiol. Mol. Biol. Reviews* 66: 506 - 577).

"Fermentação" ou "processo de fermentação" refere-se a qualquer processo de fermentação ou qualquer processo compreendendo uma etapa de fermentação. Um processo de fermentação inclui, sem limitação, processos de fermentação usados para produzir produtos de fermentação incluindo álcoois (por exemplo, arabinitol, butanol, etanol, glicerol, metanol, 1,3-propanodiol, sorbitol, e xilitol); ácidos orgânicos (por exemplo, ácido

acético, ácido acetônico, ácido adípico, ácido ascórbico, ácido cítrico, ácido 2,5-diceto-D-glucônico, ácido fórmico, ácido fumárico, ácido glucárico, ácido glucônico, ácido glucurônico, ácido glutárico, ácido 3-hidroxiopropiônico, ácido itacônico, ácido láctico, ácido málico, ácido malônico, ácido oxálico, ácido propiônico, ácido succínico, e ácido xilônico); cetonas (por exemplo, acetona); aminoácidos (por exemplo, ácido aspártico, ácido glutâmico, glicina, lisina, serina, e treonina); gases (por exemplo, metano, hidrogênio (H_2), dióxido de carbono (CO_2), e monóxido de carbono (CO)). Processos de fermentação também incluem processos de fermentação usados na indústria de álcool consumível (por exemplo, cerveja e vinho), indústria de laticínios (por exemplo, laticínios fermentados), indústria de couro, e indústria de fumo.

A proteína de fusão ou um componente da mesma tendo atividade celulolítica ou atividade hemi-celulolítica obtida de acordo com os métodos da presente invenção, por exemplo, uma proteína de fusão beta-glucosidase ou um componente da mesma, e células hospedeiras da mesma, pode ser usada na produção de monossacarídeos, dissacarídeos, e polissacarídeos, como cargas de alimentação químicas ou de fermentação a partir da biomassa para a produção de etanol, plásticos e outros produtos ou intermediários. Em particular, os polipeptídeos e as células hospedeiras da presente invenção podem ser usados para aumentar o valor de resíduos de processamento (grãos secos de destiladores, grãos gastos de cervejaria, bagaço de cana de açúcar, etc.) por degradação completa ou parcial de celulose ou hemicelulose. Para reforçar o processamento de materiais celulósicos pela preparação de enzima celulolítica em glicose, xilose, manose, galactose e arabinose, seus polímeros ou produtos derivados dos mesmos como mostrado abaixo, os polipeptídeos podem estar na forma de caldo de fermentação bruto com ou sem as células ou na forma de uma preparação de enzima semi-purificada ou purificada. O polipeptídeo pode ser uma preparação de mono-componente, uma preparação de proteína de multi-

componentes ou uma combinação de preparações de proteína multi-componentes e mono-componente. Alternativamente, uma célula hospedeira pode ser usada como uma fonte de tal polipeptídeo em um processo de fermentação com a biomassa. A célula hospedeira também pode conter genes heterólogos ou nativos que codificam proteína celulolítica, assim como outras enzimas utilizáveis no processamento de biomassa.

A presente invenção ainda refere-se a métodos para produzir uma substância orgânica, compreendendo: (a) sacarificar um material celulósico com uma quantidade eficaz de uma composição de enzima celulolítica na presença de uma quantidade eficaz de uma proteína de fusão ou componente da mesma tendo atividade celulolítica ou atividade hemi-celulolítica obtida de acordo com os presentes métodos, (b) fermentar o material celulósico sacarificado de etapa (a) com um ou mais microorganismos de fermentação; e (c) recuperar a substância orgânica da fermentação. Como indicado acima, para fins de ilustração, um polipeptídeo tendo atividade beta-glucosidase obtido de acordo com os métodos da presente invenção, por exemplo proteína de fusão de beta-glucosidase ou um componente da mesma, é usado para fins ilustrativos. O polipeptídeo tendo atividade beta-glucosidase pode estar na forma de um caldo de fermentação bruto com ou sem as células ou na forma de uma preparação de enzima semi-purificada ou purificada. A proteína beta-glucosidase pode ser uma preparação mono-componente, uma preparação de proteína multi-componentes ou uma combinação de preparações multi-componentes e mono-componente.

A substância pode ser qualquer substância derivada da fermentação. Em um aspecto preferido, a substância é um álcool. Deve-se entender que o termo "álcool" engloba uma substância que contém uma ou mais porções de hidroxila. Em um aspecto mais preferido, o álcool é arabinitol. Em outro aspecto mais preferido, o álcool é butanol. Em outro

aspecto mais preferido, o álcool é etanol. Em outro aspecto mais preferido, o álcool é glicerol. Em outro aspecto mais preferido, o álcool é metanol. Em outro aspecto mais preferido, o álcool é 1,3-propanodiol. Em outro aspecto mais preferido, o álcool é sorbitol. Em outro aspecto mais preferido, o álcool é xilitol. Ver, por exemplo, Gong, C. S., Cao, N. J., Du, J., e Tsao, G. T., 1999, Ethanol production from renewable resources, in *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heideiberg, Alemanha, 65: 207 - 241; Silveira, M. M., e Jonas, R., 2002, The biotechnological production of sorbitol, *Appl. Microbiol. Biotechnol.* 59: 400 - 408; Nigam, P., e Singh, D., 1995, Processes for fermentative production of xylitol - a sugar substitute, *Process Biochemistry* 30 (2): 117 - 124; Ezeji, T. C., Qureshi, N. e Blaschek, H. P., 2003, Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and *in situ* recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19 (6): 595 - 603.

Em outro aspecto preferido, a substância é um ácido orgânico. Em outro aspecto mais preferido, o ácido orgânico é ácido acético. Em outro aspecto mais preferido, o ácido orgânico é ácido acetônico. Em outro aspecto mais preferido, o ácido orgânico é ácido adípico. Em outro aspecto mais preferido, o ácido orgânico é ácido ascórbico. Em outro aspecto mais preferido, o ácido orgânico é ácido cítrico. Em outro aspecto mais preferido, o ácido orgânico é ácido 2,5-diceto-D-glucônico. Em outro aspecto mais preferido, o ácido orgânico é ácido fórmico. Em outro aspecto mais preferido, o ácido orgânico é ácido fumárico. Em outro aspecto mais preferido, o ácido orgânico é ácido glucárico. Em outro aspecto mais preferido, o ácido orgânico é ácido glucônico. Em outro aspecto mais preferido, o ácido orgânico é ácido glucurônico. Em outro aspecto mais preferido, o ácido orgânico é ácido glutárico. Em outro aspecto preferido, o ácido orgânico é ácido 3-hidroxipropiônico. Em outro aspecto mais preferido, o ácido orgânico é ácido

itacônico. Em outro aspecto mais preferido, o ácido orgânico é ácido láctico. Em outro aspecto mais preferido, o ácido orgânico é ácido málico. Em outro aspecto mais preferido, o ácido orgânico é ácido malônico. Em outro aspecto mais preferido, o ácido orgânico é ácido oxálico. Em outro aspecto mais preferido, o ácido orgânico é ácido propiônico. Em outro aspecto mais preferido, o ácido orgânico é ácido succínico. Em outro aspecto mais preferido, o ácido orgânico é ácido xilônico. Ver, por exemplo, Chen, R., e Lee, Y. Y., 1997, Membrane-mediated extractive fermentation for lactic acid production from cellulosic biomass, *Appl. Biochem, Biotechnol.* 63-65: 435 - 448.

Em outro aspecto preferido, a substância é uma cetona. Deve-se entender que o termo "cetona" engloba uma substância que contém uma ou mais porções de cetona. Em outro aspecto mais preferido, a cetona é acetona. Ver, por exemplo, Qureshi e Blaschek, 2003, *supra*.

Em outro aspecto preferido, a substância é um aminoácido. Em outro aspecto mais preferido, o ácido orgânico é ácido aspártico. Em outro aspecto mais preferido, o aminoácido é ácido glutâmico. Em outro aspecto mais preferido, o aminoácido é glicina. Em outro aspecto mais preferido, o aminoácido é lisina. Em outro aspecto mais preferido, o aminoácido é serina. Em outro aspecto mais preferido, o aminoácido é treonina. Ver, por exemplo, Richard, A., e Margaritis, A., 2004, Empirical modeling of batch fermentation kinetics for poly (glutamic acid) production and other microbial biopolymers, *Biotechnology and Bioengineering* 87 (4): 501 - 515.

Em outro aspecto preferido, a substância é um gás. Em outro aspecto mais preferido, o gás é metano. Em outro aspecto mais preferido, o gás é H₂. Em outro aspecto mais preferido, o gás é CO₂. Em outro aspecto mais preferido, o gás é CO. Ver, por exemplo, Kataoka, N., A. Miya, e K. Kiriya, 1997, Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria, *Water Science and*

Technology 38 (6-7): 41 - 47; e Gunaseelan V. N. in *Biomass and Bioenergy*, vol. 13 (1-2), pp. 83 - 114, 1997, Anaerobic digestion of biomass for methane production: A review.

5 Produção de uma substância a partir de material celulósico tipicamente requer quatro etapas principais. Estas quatro etapas são pré-tratamento, hidrólise enzimática, fermentação, e recuperação. Exemplificado abaixo é um processo para produzir etanol, mas deve-se entender que processos similares podem ser usados para produzir outras substâncias, por exemplo, as substâncias descritas acima.

10 Pré-tratamento. No pré-tratamento ou etapa de pré-hidrólise, o material celulósico é aquecido para romper a lignina e estrutura de carboidrato, solubilizar a maior parte da hemicelulose, e tornar a fração de celulose acessível para enzimas celulolíticas. O aquecimento é realizado ou
15 diretamente com vapor ou em suspensão onde um catalisador também pode ser adicionado no material para acelerar as reações. Os catalisadores incluem ácidos fortes, como ácido sulfúrico e SO₂, ou alcalino, como hidróxido de sódio. O propósito do estágio de pré-tratamento é para facilitar a penetração das enzimas e microorganismos. A biomassa celulósica também pode ser submetida a um pré-tratamento de explosão a vapor hidrotérmico (ver pedido
20 de patente 20020164730). No entanto, entende-se que na prática dos métodos da presente invenção, qualquer pré-tratamento pode ser usado empregando pré-tratamento térmico, químico e/ou mecânico.

25 Sacarificação. Na etapa de hidrólise enzimática, também conhecida como sacarificação, as enzimas como descrito aqui são adicionadas no material prétratado para converter a fração de celulose para glicose e/ou outros açúcares. A sacarificação geralmente é realizada em reatores de tanques agitados ou fermentadores sob condições controladas de pH, temperatura, e mistura. Uma etapa de sacarificação pode durar 200 horas. A sacarificação pode ser realizada em temperaturas de cerca de 30 °C a cerca de

65 °C, em particular, em torno de 50 °C, e a um pH na faixa entre cerca de 4 a cerca de 5, especialmente em torno de pH 4,5. Para produzir glicose que pode ser metabolizada por levedura, a hidrólise é tipicamente realizada na presença de uma beta-glucosidase.

5 Fermentação. Na etapa de fermentação, açúcares, liberados do material celulósico como um resultado das etapas de hidrólise enzimática e pré-tratamento, são fermentados em etanol por um organismo de fermentação, como levedura. A fermentação também pode ser realizada simultaneamente com a hidrólise enzimática no mesmo vaso, novamente sob condições
10 controladas de pH, temperatura, e mistura. Quando a sacarificação e fermentação são realizadas simultaneamente no mesmo vaso, o processo geralmente é chamado sacarificação e fermentação simultâneas ou SSF.

Qualquer material celulósico ou matéria prima apropriado pode ser usado em etapa de fermentação na prática da presente invenção. O
15 material geralmente é selecionado com base no produto de fermentação desejado, isto é, a substância a ser obtida da fermentação, e o processo empregado, como é bem conhecido na arte. Exemplos de substratos apropriados para usar nos métodos da presente invenção incluem materiais contendo celulose, como resíduos de madeira ou planta ou açúcares de baixo
20 peso molecular DP1-3 obtidos a partir de material celulósico processado que pode ser metabolizado pelo microorganismo de fermentação, e que pode ser fornecido por adição direta ao meio de fermentação.

O termo "meio de fermentação" será entendido como fazendo referência a um meio antes do(s) microorganismo (s) de fermentação ser
25 adicionado(s), como, um meio resultante de um processo de sacarificação, assim como um meio usado em um processo de sacarificação e fermentação simultâneas (SSF).

"Microorganismo de fermentação" refere-se a qualquer microorganismo apropriado para uso em um processo de fermentação

desejado. Os microorganismos de fermentação apropriados de acordo com a invenção são capazes de fermentar, isto é, converter, açúcares, como glicose, xilose, arabinose, manose, galactose, ou oligossacarídeos diretamente ou indiretamente no produto de fermentação desejado. Exemplos de

5 microorganismos de fermentação incluem organismos de fungos, como leveduras. As leveduras preferidas incluem cepas de *Saccharomyces spp.*, e em particular, *Saccharomyces cerevisiae*. A levedura comercialmente disponível inclui, por exemplo, Red Star^{®/TM}/Lesaffre Ethanol Red (disponível de Red Star/Lesaffre, USA) FALI (disponível de Fleischmann's

10 Yeast, uma divisão de Burns Philp Food Inc., USA), SUPERSTART (disponível de Alltech), GERT STRAND (disponível de Gert Strand AB, Suécia) e FERMIOL (disponível de DSM Specialties).

Em um aspecto preferido, a levedura é *Saccharomyces spp.*

Em um aspecto mais preferido, a levedura é *Saccharomyces cerevisiae*. Em

15 outro aspecto mais preferido, a levedura é *Saccharomyces distaticus*. Em outro aspecto mais preferido, a levedura é *Saccharomyces uvarium*. Em outro aspecto preferido, a levedura é *Kluyveromyces*. Em outro aspecto mais preferido, a levedura é *Kluyveromyces marxianus*. Em outro aspecto mais preferido, a levedura é *Kluyveromyces fragilis*. Em outro aspecto preferido, a

20 levedura é *Candida*. Em outro aspecto mais preferido, a levedura é *Candida pseudotropicalis*. Em outro aspecto mais preferido, a levedura é *Candida brassicae*. Em outro aspecto preferido, a levedura é uma *Clavispora*. Em outro aspecto mais preferido, a levedura é *Clavispora lusitaniae*. Em outro aspecto mais preferido, a levedura é *Clavispora opuntiae*. Em outro aspecto

25 preferido, a levedura é *Pachysolen*. Em outro aspecto mais preferido, a levedura é *Pachysolen tannophilus*. Em outro aspecto preferido, a levedura é *Bretannomyces*. Em outro aspecto mais preferido, a levedura é *Bretannomyces clausenii* (Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman,

C. E., ed., Taylor & Francis, Washington, DC, 179 - 212).

Bactérias que podem eficientemente fermentar glicose em etanol incluem, por exemplo, *Zymomonas mobilis* e *Clostridium thermocellum* (Philippidis, 1996, *supra*).

5 É bem conhecido na técnica que os organismos descritos acima também podem ser usados para produzir outras substâncias, como descrito aqui.

A clonagem de genes heterólogos em *Saccharomyces cerevisiae* (Chen, Z., Ho, N. W. Y., 1993, Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, *Appl. Biochem. Biotechnol.* 39 - 40: 135 - 147; Ho, N. W. Y., Chen, Z, Brainard, A. P., 1998, Genetically engineering *Saccharomyces* yeast capable of effectively cofermenting glicose and xylose, *Appl. Environ. Microbiol.* 64: 1852 - 1859), ou em bactérias como *Escherichia coli* (Beall, D. S., Qhta, K., Ingram, L. O., 1991, Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*, *Biotech. Bioeng.* 38: 296 - 303), *Klebsiella oxytoca* (Ingram, L. O., Gomes, P. F., Lai, X., Moniruzzaman, M., Wood, B. E., Yomano, L. P., York, S. W., 1998, Metabolic engineering of bacteria for ethanol production, *Biotechnol. Bioeng.* 58: 204 - 214), e *Zymomonas mobilis* (Zhang, M., Eddy, C., Deanda, K., Finkelstein, M., e Picataggio, S., 1995, Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobitis*, *Science* 267: 240 - 243; Deanda, K., Zhang, M., Eddy, C., e Picataggio, S., 1996, Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering, *Appl. Environ. Microbiol.*, 62: 4465 - 4470) têm levado à construção de organismos capazes de converter hexoses e pentoses em etanol (cofermentação).

Levedura ou outro microorganismo tipicamente é adicionado à celulose ou hidrolisado degradado e a fermentação prossegue durante cerca de

24 a cerca de 96 horas, como cerca de 35 e a cerca de 60 horas. A temperatura está tipicamente entre cerca de 26 °C a cerca de 40 °C, em particular de cerca de 32 °C, a cerca de pH 3 a cerca de pH 6, em particular em torno de pH 4-5.

5 Em aspecto preferido, levedura ou outro microorganismo é aplicado na celulose degradada ou hidrolisado e a fermentação é realizada durante cerca de 24 a cerca de 96 horas, como tipicamente 35-60 horas. Em um aspecto preferido, a temperatura geralmente é entre cerca de 26 a cerca de 40 °C, em particular de cerca de 32 °C, e o pH geralmente é de cerca de pH 3 a cerca de pH 6, preferivelmente em torno de pH 4-5. Levedura ou outro
10 microorganismo é preferivelmente aplicado em quantidades de aproximadamente 10^5 a 10^{12} , preferivelmente de aproximadamente 10^7 a 10^{10} , especialmente aproximadamente 5×10^7 contagem viável por ml de caldo de fermentação. Durante uma fase de produção de etanol, a contagem de célula de levedura deve preferivelmente estar na faixa de aproximadamente 10^7 a
15 10^{10} , especialmente em torno de aproximadamente 2×10^8 . Outra diretriz com relação ao uso de levedura para fermentação pode ser encontrada em, por exemplo, "The Alcohol Textbook" (Editores K. Jacques, T. P. Lyons e D. R. Kelsall, Nottingham University Press, Reino Unido 1999), que é incorporado aqui por referência.

20 O processo na técnica mais amplamente usado é o processo de sacarificação e fermentação simultâneas (SSF) onde não se tem um estágio de manutenção para a sacarificação, significando que a levedura e a enzima são adicionadas juntas.

25 Para a produção de etanol, após a fermentação, a massa é destilada para extrair o etanol. O etanol obtido de acordo com o processo da invenção pode ser usado como, por exemplo, etanol combustível; etanol para bebidas, isto é, álcool neutro potável, ou etanol industrial.

Um estimulador de fermentação pode ser usado em combinação com qualquer um dos processos enzimáticos descritos aqui para

ainda melhorar o processo de fermentação, e em particular, o desempenho do microorganismo de fermentação, como, melhora da taxa e rendimento de etanol. Um "estimulador de fermentação" refere-se a estimuladores para crescimento dos microorganismos de fermentação, em particular, levedura.

- 5 Estimuladores de fermentação preferidos para crescimento incluem vitaminas e minerais. Exemplos de vitaminas incluem multivitaminas, biotina, pantotenato, ácido nicotínico, meso-inositol, tiamina, piridoxina, ácido para-aminobenzóico, ácido fólico, riboflavina, e vitaminas A, B, C, D, e E. Ver, por exemplo, Alfenore *et al.*, Improving ethanol production and viability of
- 10 *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag (2002), que é incorporado aqui por referência. Exemplos de minerais incluem minerais e sais minerais que podem suprir nutrientes compreendendo P, K, Mg, S, Ca, Fe, Zn, Mn, e Cu.

Recuperação. O álcool é separado do material celulósico fermentado e purificado por métodos convencionais de destilação. Etanol com 15 uma pureza de até cerca de 96% em volume de etanol pode ser obtido, que pode ser usado como, por exemplo, etanol combustível, etanol para bebidas, isto é, álcool neutro potável, ou etanol industrial.

Para outras substâncias, qualquer método conhecido na técnica 20 pode ser usado incluindo, mas não limitado a, cromatografia (por exemplo, troca iônica, afinidade, hidrofóbico, cromatofocalização, e exclusão por tamanho), procedimentos eletroforéticos (por exemplo, focalização isoelétrica preparativa), solubilidade diferencial (por exemplo, precipitação com sulfato de amônio), SDS-PAGE, destilação, ou extração.

25 Nos métodos da presente invenção, além de beta-glucosidase, a preparação de enzima celulolítica e polipeptídeo(s) melhorador(es) celulolítico(s) podem ser suplementados por uma ou mais atividades de enzima adicionais para melhorar a degradação do material celulósico. As enzimas adicionais preferidas são hemicelulases, esterases (por exemplo,

lipases, fosfolipases, e/ou cutinases), proteases, lacases, peroxidases, ou misturas das mesmas.

Nos métodos da presente invenção, a (s) enzima (s) adicional (ais) pode (m) ser adicionada (s) antes ou durante a fermentação, incluindo
5 durante ou após a propagação do (s) microorganismo (s) de fermentação.

As enzimas referidas aqui podem ser derivadas ou obtidas de qualquer origem apropriada, incluindo, origem de bactérias, fungos, leveduras ou mamíferos. O termo "obtido" significa aqui que a enzima pode ter sido isolada de um organismo que naturalmente produz a enzima como uma
10 enzima nativa. O termo "obtido" também significa aqui que a enzima pode ter sido produzida recombinantemente em um organismo hospedeiro, em que a enzima recombinantemente produzida é ou nativa ou estranha para o organismo hospedeiro ou tem uma sequência de aminoácido modificada, por exemplo, tendo um ou mais aminoácidos que são deletados, inseridos e/ou
15 substituídos, isto é, uma enzima recombinantemente produzida que é um mutante e/ou um fragmento de uma sequência de aminoácido nativa ou uma enzima produzida por processos de embaralhamento de ácido nucléico conhecidos na arte. Estão englobadas dentro do significado de uma enzima nativa as variantes naturais e dentro do significado de uma enzima estranha as
20 variantes recombinantemente obtidas, como por mutagênese sítio-dirigido ou embaralhamento.

As enzimas também podem ser purificadas. O termo "purificado" como usado aqui cobre as enzimas livres de outros componentes do organismo das quais ele é derivado. O termo "purificado" também cobre as
25 enzimas livres de componentes do organismo nativo a partir das quais ele é obtido. As enzimas podem ser purificadas, com somente quantidades menores de outras proteínas estando presentes. A expressão "outras proteínas" refere-se em particular a outras enzimas. O termo "purificado" como usado aqui também refere-se à remoção de outros componentes, particularmente outras

proteínas e o mais particularmente outras enzimas presentes na célula de origem da enzima da invenção. A enzima pode ser "substancialmente pura," isto é, livre de outros componentes do organismo em que é produzida, isto é, por exemplo, um organismo hospedeiro para enzimas recombinantemente produzidas.

As enzimas usadas na presente invenção podem estar em qualquer forma apropriada para uso nos processos descritos aqui, como, por exemplo, um caldo de fermentação bruto com ou sem células, um pó seco ou granulado, um granulado não pulverulento, um líquido, um líquido estabilizado, ou uma enzima protegida. Granulados podem ser produzidos, por exemplo, como descrito na patente U.S. 4.106.991 e 4.661.452, e opcionalmente podem ser revestidos por processos conhecidos na arte. Preparações de enzima líquida podem, por exemplo, ser estabilizadas por adição de estabilizadores como um açúcar, um álcool de açúcar ou outro poliol, e/ou ácido láctico ou outro ácido orgânico de acordo com o processo estabelecido. As enzimas protegidas podem ser preparadas de acordo com o processo descrito em EP 238.216.

Composições detergentes

Os métodos da presente invenção são particularmente utilizáveis para melhorar a secreção de polipeptídeos em quantidades comercialmente importantes para uso em composições detergentes. Estes polipeptídeos incluem, mas não são limitados a proteases, enzimas celulolíticas, amilases e peroxidases, ou qualquer outra enzima ou proteína biológica utilizável para a indústria de detergentes.

A composição detergente da presente invenção pode ser por exemplo, formulada como uma composição detergente para lavagem manual ou em máquina, incluindo uma composição aditiva para lavagem de roupas, apropriada para pré-tratamento de tecidos manchados e uma composição amaciante de tecido adicionada no enxágüe, ou formulada como uma

composição detergente para uso em operações de limpeza de superfícies duras domésticas em geral ou formuladas para operações de lavagem de louça manual ou em máquina.

Em um aspecto específico, a presente invenção provê um
5 aditivo detergente compreendendo um polipeptídeo tendo atividade biológica, (por exemplo, proteína de fusão, um componente da mesma, ou combinações) obtidos de acordo com a presente invenção. O aditivo detergente assim como a composição detergente podem compreender uma ou mais de outras enzimas como uma protease, lipase, cutinase, uma amilase, carbohidrase celulase,
10 pectinase, mannanase, arabinase, galactanase, xilanase, oxidase, por exemplo, uma laccase, e/ou peroxidase.

Em geral, as propriedades dos componentes enzimáticos devem ser compatíveis com o detergente selecionado (isto é, ótimo pH, compatibilidade com outros ingredientes enzimáticos e não enzimáticos, etc),
15 e os componentes enzimáticos devem estar presentes em quantidades eficazes.

Proteases - As proteases apropriadas incluem as de origem animal, vegetal ou microbiana. A origem microbiana é preferida. Os mutantes engenheirados por proteínas ou quimicamente modificados são incluídos. A protease pode ser serina protease, ou uma metalo protease, preferivelmente
20 uma protease microbiana alcalina ou protease de tipo tripsina. Os exemplos de proteases alcalinas são as subtilisinas, especialmente as derivadas de *Bacillus*, por exemplo, subtilisina Novo, subtilisina Carlsberg, subtilisina 309, subtilisina 147 e subtilisina 168 (descritas em WO 89/06279). Exemplos de proteases de tipo tripsina são tripsina (por exemplo, de origem de porcos ou
25 bovinos) e a protease de *Fusarium*, como descrito em WO 89/06270 e WO 94/25583.

Exemplos de proteases úteis são as variantes, como descrito em WO 92/19729, WO35 98/20115, WO 98/20116, e WO 98/34946, especialmente as variantes com substituições em uma ou mais das seguintes

posições 27, 36, 57, 76, 87, 97, 101, 104, 120, 123, 167, 170, 194, 206, 218, 222, 224, 235 e 274.

As enzimas de protease comercialmente disponíveis preferidas incluem ALCALASE™, SAVINASE™, PRIMASE™, DURALASE™, ESPERASE™, e KANNASE™ (NOVOZYMES A/S), MAXATASE™, MAXACAL™, MAXAPEM™, PROPERASE™, PURAFECT™, PURAFECT OXP™, FN2™, e FN3™ (Genencor International Inc).

Lipases. As lipases apropriadas incluem as de origem bacteriana ou fúngica. Os mutantes engenheirados por proteínas ou quimicamente modificados são incluídos. Os exemplos de lipases úteis incluem lipases de *Humicola* (sinônimo *Thermomyces*) por exemplo, de *H. lanuginosa* (*T lanuginosus*) como descrito em EP. 258 068 e EP. 305 216 ou de *H. insolens* como descrito em WO 96/13580, uma lipase de *Pseudomonas*, por exemplo, de *P. alcaligenes* ou *P. pseudoalcaligenes* (EP. 218 272), *P. cepacia* (EP. 331 376), *P. stutzeri* (GB 1.372.034), *P. fluorescens*, *Pseudomonas* sp cepa SD 705 (WO 95/06720 e WO 96/27002), *P. wisconsinensis* (WO 96/12012), uma lipase de *Bacillus*, por exemplo, de *B. subtilis* (Dartois et al, 1993, *Biochemica et Biophysica Acta*, 1131, 253-360), *B. stearothermophilus* (JP. 64/744992) ou *B. pumilus* (WO 91/16422).

Outros exemplos são variantes de lipase como descrito em WO 92/05249, WO 94/01541, EP 407 225, EP 260 105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783 WO 95/22615, WO 97/04079 e WO 97/07202.

As lipases comercialmente disponíveis preferidas incluem LIPOLASE™, LIPEX™, e Lipolase ULTRA™ (Novozymes A/S).

Amilases. As amilases apropriadas (α e/ou β) incluem as de origem bacteriana ou fúngica. Os mutantes engenheirados por proteínas ou quimicamente modificados são incluídos. As amilases incluem, por exemplo,

α -amilases obtidas a partir de *Bacillus*, por exemplo, uma cepa

especial de *Bacillus licheniformis*, como descrito em maiores detalhes em GB 1.296.839.

Exemplos de amilases utilizáveis são as variantes como descrito em WO 94/02597, WO 94/18314, WO 96/23873, e WO 97/43424, especialmente as variantes com substituições em uma ou mais das seguintes posições: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408, e 444.

As amilases comercialmente disponíveis são DURAMYL™, TERMAMYL™, FUNGAMYL™ e BAN™ (Novozymes A/S), Rapidase™ e Purastar™ (de Genencor International Inc).

Celulases As celulases apropriadas incluem as de origem bacteriana ou fúngica. Os mutantes engenheirados por proteínas ou quimicamente modificados são incluídos. As celulases apropriadas incluem celulases dos gêneros *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, ou *Trichoderma* por exemplo, celulases fúngicas produzidas a partir de *Humicola insolens*, *Myceliophthora thermophila* e *Fusarium oxysporum* descritas em Patente US No 4.435.307, Patente US No 5.648.263, Patente US No 5.691.178, Patente US No 5.776.757 e WO 89/09259.

As celulases especialmente apropriadas são as celulases neutras ou alcalinas tendo benefícios de cuidado da cor. Os exemplos de tais celulases são celulases como descrito em EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Outros exemplos são variantes de celulase como as descritas em WO 94/07998, EP 0 531 315, Patente US No 5.457.046, Patente US No 5.686.593, Patente US No 5.763.254, WO 95/24471, WO 98/12307 e PCT/DK98/00299.

As celulases comercialmente disponíveis incluem CELLUCLAST®, CELLUZYME™, e CAREZYME™ (Novozymes A/S), Clazinase™, e Puradax HA™ (Genencor International Inc), e KAC-500(B)™

(Kao Corporation).

Peroxidasas/Oxidases. Apropriadas peroxidases/oxidases incluem as de origem de plantas, bacterianas ou fúngicas. Os mutantes engenheirados por proteínas ou quimicamente modificados são incluídos. Os
5 exemplos de peroxidases utilizáveis incluem peroxidases de *Coprinus*, por exemplo, de *C. cinereus*, e variantes das mesmas, como descrito em WO 93/24618, WO 95/10602, e WO 98/15257.

As peroxidases comercialmente disponíveis incluem GUARDZYME™ (Novozymes A/S).

10 Os componente(s) enzimático(s) podem ser incluídos em uma composição detergente por adição de aditivos separados contendo uma ou mais enzimas, ou por adição de um aditivo combinado compreendendo todas estas enzimas. Um aditivo detergente da presente invenção, isto é, um aditivo separado ou aditivo combinado, pode ser formulado, por exemplo, como um
15 granulado, líquido, suspensão, etc. As formulações aditivas detergentes preferidas são granulados, particularmente granulados não pulverulentos, líquidos, particularmente líquidos estabilizados ou suspensões.

Granulados não pulverulentos podem ser produzidos, por exemplo, como descrito na patente U.S. 4.106.991 e 4.661.452, e
20 opcionalmente podem ser revestidos por processos conhecidos na arte. Os exemplos de materiais de revestimento ceroso são produtos de poli (óxido de etileno) (polietileno glicol PEG) com pesos molares médios de 1.000 a 20.000; nonilfenóis etoxilados tendo de 16 a 50 unidades de óxido de etileno, álcoois graxos etoxilados, em que o álcool contém de 12 a 20 átomos de
25 carbono, e em que se tem 15 a 80 unidades de óxido de etileno; álcoois graxos, ácidos graxos; e mono- e di- e triglicerídeos de ácidos graxos. Exemplos de materiais de revestimento de formação de películas apropriados para aplicação por técnicas de leito fluido são dados em GB 1483591. Preparações de enzima líquida podem, por exemplo, ser estabilizadas por

adição de um poliol como propileno glicol, um açúcar, um álcool de açúcar ou ácido bórico, de acordo com os processos estabelecidos. As enzimas protegidas podem ser preparadas de acordo com o processo descrito em EP 238.216.

5 A composição detergente da presente invenção pode estar em qualquer forma conveniente, por exemplo uma barra, um tablete, um pó, um grânulo, uma pasta ou um líquido. Um detergente líquido pode ser aquoso, tipicamente contendo até 70% de água e 0-30% de solvente orgânico, ou não aquoso.

10 A composição detergente compreende um ou mais tensoativos, que podem ser não iônicos incluindo semi-polar e/ou aniônico e/ou catiônica e/ou zwitteriônico. Os tensoativos estão tipicamente presentes em um nível de 0,1 a 60 % em peso.

15 Quando incluído ai, o detergente irá geralmente conter de cerca de 1% a cerca de 40% de um tensoativo aniônico como alquilbenzenossulfato linear, alfa-olefina sulfonato, alquil sulfato (sulfato de álcool graxo), etoxissulfato de álcool, alcanossulfonato secundário, éster metílico de alfa-sulfo ácido graxo, ácido alquil ou alquenil succínico ou sabão.

20 Quando incluído ai, o detergente irá geralmente conter de cerca de 0,2% a cerca de cerca de 40% de um tensoativo não iônico como etoxilato de álcool, etoxilato de nonilfenol, alquilpoliglicosídeo, óxido de alquildimetilamina etoxilada, monoetanolamida de ácido graxo etoxilada,

25 Monoetanolamida de ácido graxo, polihidroxi alquil amida de ácido graxo, ou derivado de N-acil N-alquila de glucosamina ("glucamidas").

O detergente pode conter de 0-65 % de um reforçador de detergente ou agente complexantes, como zeólito, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilenodiaminatetraacético, ácido dietilenotriaminapentaacético, ácido alquil-

ou alquenilsuccínico, silicatos solúveis ou silicatos em camadas (por exemplo, SKS-6 de Hoechst).

5 O detergente pode compreender um ou mais polímeros. Os exemplos são carboximetilcelulose, polivinilpirrolidona), polietileno glicol), poli(álcool vinílico), poli(vinilpiridina-N- óxido)), poli(vinilimidazol), policarboxilatos como poliacrilatos, copolímeros de ácido maleico/ acrílico, e copolímeros de metacrilato de laurila / ácido acrílico.

10 O detergente pode conter um sistema alvejante que pode compreender uma fonte de H_2O_2 como perborato ou percarbonato que pode ser combinada com um ativador alvejante formador de perácido, como tetraacetileno diamina ou nonanoiloxibenzenossulfonato. Alternativamente, o sistema alvejante pode compreender peroxiácidos de, por exemplo, o tipo amida, imida ou sulfona.

15 O(s) componente(s) enzimático(s) da composição detergente da presente invenção pode ser estabilizado usando agentes estabilizantes convencionais, por exemplo um poliol como propileno glicol ou glicerol, um açúcar ou álcool de açúcar, ácido bórico ou derivado de ácido bórico, por exemplo um éster de borato aromático, ou derivado de ácido fenil borônico como ácido 4- formilfenil borônico, e a composição pode ser formulada como
20 descrito em, por exemplo, WO 30 92/19709 e WO 92/19708.

O detergente pode também conter outros ingredientes detergentes convencionais, como condicionadores de tecido incluindo argilas, reforçadores de espuma, supressores de espuma, agentes anti-corrosão, agentes de colocar em suspensão as sujeiras, agentes anti-redeposição de
25 sujeira, corantes, bactericidas, brilhantadores ópticos, hidrótroos, inibidores de manchas, ou perfumes.

Nas composições detergentes, qualquer componente enzimático, particularmente os polipeptídeos tendo atividade biológica da presente invenção podem ser adicionados em uma quantidade correspondendo

a 0,01- 100 mg de proteína de enzima por litro de licor de lavagem, preferivelmente 0,05 - 5 mg de proteína de enzima por litro de licor de lavagem, particularmente 0,1 - 1 mg de proteína de enzima por litro de licor de lavagem.

5 Os polipeptídeos tendo atividade biológica da presente invenção podem ser incorporados adicionalmente nas formulações detergentes descritas em WO 97/07202, que se incorpora aqui por referência.

Outros usos

10 Polipeptídeos tendo atividade biológica (por exemplo, proteína de fusão, um componente da mesma ou combinações das mesmas) obtidos de acordo com a presente invenção também podem ser usados em combinação com outras glicohidrolases e enzimas relacionadas, com aqui descrito, no tratamento de têxteis como agentes de biopolimento ou para redução de felpas, bolinhas, modificação da textura e lavagem com pedra (N. K. Lange, 15 in P. Suominen, T. Reimkainen (Eds), *Trichoderma reesei Cellulases and Other Hidrolases*, Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, 1993, pp 263-272). Além disso, os polipeptídeos descrito também podem ser usados em combinação com outras glicohidrolases e 20 enzimas relacionadas, como aqui descrito, em processamento de madeira para formação de biopasta ou remoção de cascas, fabricação de papel para modificação de fibras, branqueamento, e redução de custos de energias de refinaria, tratamento de água branca, importantes para o reciclo de água de refugo, reciclagem de fibras lignocelulósicas, como remoção de tinta e processamento secundário da fibra, e utilização de resíduos de madeira (S. D. 25 Mansfield e A. R. Esteghlalian em S D, Mansfield e J. N. Saddler (Eds), *Applications of Enzimas to lignocelulosics*, ACS Symposium Series 855, Washington, D C, 2003 pp 2-29)

A presente invenção é ainda descrita nos seguintes exemplos que não devem ser construídos como limitando o escopo da invenção.

Exemplos

Produtos químicos usados como tampões e substratos foram produtos comerciais de pelo menos tipo reagente.

Sequenciamento de DNA

5 O Sequenciamento de DNA foi realizado usando um analisador genético modelo 3130X de Applied Biosystems (Applied Biosystems, Foster City, CA, USA) usando química de terminador de corante (Giesecke *et al.*, 1992, *Journal of Virol. Methods* 38: 47 - 60). As sequências foram montadas usando phred/phrap/consed (University of Washington, Seattle, WA, USA) com iniciadores de sequência específicos.

Cepa

RutC30 de *Trichoderma reesei* (ATCC 56765; Montenecourt e Eveleigh, 1979, *Adv. Chem. Ser.* 181: 289 - 301) foi derivado de *Trichoderma reesei* Qm6A (ATCC 13631; Mandels e Reese, 1957, *J. Bacteriol.* 73: 269 - 15 278). RutC30 de *Trichoderma reesei* e cepa *Aspergillus oryzae* Jal355 (WO 02/062973) foram usados para expressão da proteína de fusão beta-glucosidase.

Meios de cultura e soluções

Meio YP foi composto por litro de 10 g de extrato de levedura e 20 g de bacto triptona.

Meio induzindo celulase foi composto por litro de 20 g de celulose, 10 g de sólidos de licor de milho, 1,45 g de $(\text{NH}_4)_2\text{SO}_4$, 2,08 g de KH_2PO_4 , 0,28 g de CaCl_2 , 0,42 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, e 0,42 ml de solução de metais trace.

25 Solução de metais de traço foi composta por litro de 216 g de $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 58 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 27 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 10 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 24 g de H_3BO_3 , e 336 g de ácido cítrico.

STC foi composto de 1 M de sorbitol, 10 mM de CaCl_2 , e 10 mM de Tris-HCl, pH 7,5.

Placas COVE foram compostas por litro de 342 g de sacarose, 10 ml de solução de sais COVE, 10 ml de 1 M de acetamida, 10 ml de 1,5 M de CsCl, e 25 g de ágar Noble.

Solução de sais COVE foi composta por litro de 26 g de KCl, 26 g de MgSO₄, 76 g de KH₂PO₄, e 50 ml de solução de metais trace COVE.

Solução de metais de traço COVE foi composta por litro de 0,04 g de Na₂B₄O₇·10H₂O, 0,4 g de CuSO₄·5H₂O, 1,2 g de FeSO₄·7H₂O, 0,7 g de MnSO₄·H₂O, 0,8 g de Na₂MoO₂·H₂O, e 10 g de ZnSO₄·7H₂O.

Placas COVE2 foram compostas por litro de 30 g de sacarose, 20 ml de solução de sais COVE, 25 g de ágar Noble, e 10 ml de 1 M de acetamida.

Placas PDA foram compostas por litro de 39 gramas de ágar dextrose de batata.

O meio LB foi composto por litro de 10 g de tripton, 5 g de extrato de levedura, 5 g de cloreto de sódio.

Placas 2X YT foram compostas por litro de 10 g de tripton, 5 g de extrato de levedura, 5 g de cloreto de sódio, e 15 g de ágar Bacto.

O meio MDU2BP foi composto por litro de 45 g de maltose, 1 g de MgSO₄·7H₂O, 1 g de NaCl, 2 g de K₂HSO₄, 12 g de KH₂PO₄, 2 g de uréia, e 500 µl de solução de metais trace AMG, o pH foi ajustado a 5,0 e então o filtro esterilizado com uma unidade de filtração de 0,22 µm.

Solução de metais de traço AMG foi composta por litro de 14,3 g de ZnSO₄·7H₂O, 2,5 g de CuSO₄·5H₂O, 0,5 g de NiCl₂·6H₂O, 13,8 g de FeSO₄·7H₂O, 85 g de MnSO₄·7H₂O, e 3 g de ácido cítrico.

Placas de meio mínimo foram compostas por litro de 6 g de NaNO₃, 0,52 g de KCl, 1,52 g de KH₂PO₄, 1 ml de solução de metais trace COVE, 20 g de ágar Noble, 20 ml de 50% de glicose, 2,5 ml de 20% de MgSO₄·7H₂O, e 20 ml de solução de carga biotina.

Solução de carga biotina foi composta por litro de 0,2 g de

biotina.

O meio SOC foi composto de 2% de triptona, 0,5% de extrato de levedura, 10 mM de NaCl, 2,5 mM de KCl, 10 mM de MgCl₂, e 10 mM de MgSO₄, seguido por adição de glicose de filtro esterilizado a 20 mM após autoclavagem.

Exemplo 1: Construção de vetor de expressão pMJ04

Vetor de expressão pMJ04 foi construído por amplificação PCR do gene celobiohidrolase 1 de *Trichoderma reesei* (*cbh1*, CEL7A) terminador de DNA genômico RutC30 de *Trichoderma reesei* usando iniciadores 993429 (anti-sentido) e 993428 (sentido) mostrados abaixo. O iniciador anti-sentido foi engenheirado para ter um sítio *Pac* I na extremidade 5' e um sítio *Spe* I na extremidade 3' do iniciador anti-sentido.

Iniciador 993429 (anti-sentido):

5'-AACGTTAATTAAGGAATCGTTTTGTGTTT-3' (SEQ ID NO:33)

Iniciador 993428 (sentido):

5'-AGTACTAGTAGCTCCGTGGCGAAAGCCTG-3' (SEQ ID NO: 34)

DNA genômico RutC30 de *Trichoderma reesei* foi isolado usando um kit DNEASY® Plant Maxi (QIAGEN Inc., Valencia, CA, USA).

As reações de amplificação (50 µl) foram compostas de tampão de reação 1X ThermoPol (New England Biolabs, Beverly, MA USA), 0,3 mM de dNTPs, 100 ng de DNA genômico RutC30 de *Trichoderma reesei*, 0,3 µM de iniciador 993429, 0,3 µM de iniciador 993428, e 2 unidades de Vent DNA polimerase (New England Biolabs, Beverly, MA, USA). As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 (Eppendorf Scientific, Inc., Westbury, NY, USA) programado para 5 ciclos cada durante 30 segundos a 94 °C, 30 segundos a 50 °C, e 60 segundos a 72 °C, seguido por 25 ciclos cada durante 30 segundos a 94 °C, 30 segundos a 65 °C, e 120 segundos a 72 °C (extensão final de 5 minutos). Os produtos de

reação foram isolados em um gel agarose a 1,0% usando 40 mM Tris base-20 mM de acetato de sódio-1 mM dissódico EDTA (TAE) tampão onde uma banda de produto de 229 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® (QIAGEN Inc., Valencia, CA, USA) de acordo com as instruções do fabricante.

O fragmento PCR resultante foi digerido com *Pac* I e *Spe* I e ligado em pAILo1 (WO 05/067531) digerido com as mesmas enzimas de restrição usando um kit de ligação rápido (Roche, Indianápolis, IN, USA), para gerar pMJ04 (figura 1).

10 Exemplo 2: Construção de pCaHj568

Plasmídeo pCaHj568 foi construído de pCaHj170 (patente US no. 5.763.254) e pMT2188. Plasmídeo pCaHj170 compreende a região de codificação de comprimento completo de endoglucanase V de *Humicola insolens* (CEL45A) (SEQ ID NO:1, que codifica a sequência de aminoácido de SEQ ID NO: 2). Construção de pMT2188 foi iniciada por amplificar PCR a origem de replicação de pUC19 de pCaHj483 (WO 98/00529) usando iniciadores 142779 e 142780 mostrados abaixo. Iniciador 142780 introduz um sítio *Bbu* I no fragmento PCR.

142779:

20 5'-TTGAATTGAAAATAGATTGATTTAAACTTC-3' (SEQ ID NO: 35)

142780:

5'-TTGCATGCGTAATCATGGTCATAGC-3' (SEQ ID NO: 36)

Um sistema EXPAND® PCR (Roche Molecular Biochemicals, Basel, Suíça) foi usado seguindo as instruções do fabricante para esta amplificação. Produtos PCR foram separados em um gel de agarose e um fragmento de 1160 bp foi isolado e purificado usando um Kit Spin de extração de gel Jetquick (Genomed, Wielandstr, Alemanha). O gene URA3 foi amplificado do vetor de clonagem geral de *Saccharomyces cerevisiae* pYES2 (Invitrogen, Carlsbad, CA, USA) usando iniciadores 140288 e 142778

mostrados abaixo usando um sistema EXPAND® PCR. Iniciador 140288 introduziu um sítio *Eco* RI no fragmento PCR.

5'-TTGAATTCATGGGTAATAACTGATAT-3' (SEQ ID NO: 37)
142778:

5 5'-AAATCAATCTATTTTCAATTCAATTCATCATT-3' (SEQ ID NO: 38)

Produtos PCR foram separados em um gel de agarose e um fragmento 1126 bp foi isolado e purificado usando um Kit Spin de extração de gel Jetquick.

Os dois fragmentos PCR foram fusionados por misturação e amplificados usando iniciadores 142780 e 140288 mostrados acima pelo método de emenda de sobreposição (Horton *et al.*, 1989, *Gene* 77: 61 - 68).
10 Produtos PCR foram separados em um gel de agarose e um fragmento 2263 bp foi isolado e purificado usando um Kit Spin de extração de gel Jetquick.

O fragmento resultante foi digerido com *Eco* RI e *Bbu* I e
15 ligado usando protocolos padrões para o fragmento maior de pCaHj483 digerido com as mesmas enzimas de restrição. A mistura de ligação foi transformada em cepa *E. coli* DB6507 *pyrF*-negativa (ATCC 35673) tornada competente pelo método de Mandel e Higa, 1970, *J. Mol. Biol.* 45: 154. Transformantes foram selecionados em meio sólido M9 (Sambrook *et al.*,
20 1989, *Molecular Cloning, A Laboratory Manual*, 2ª. edição, Cold Spring Harbor Laboratory Press) suplementado por litro com 1 g de casaminoácidos, 500 µg de tiamina, e 10 mg de canamicina. Um plasmídeo de um transformante foi isolado e designado pCaHj527 (figura 2).

O promotor NA2-*tpi* presente em pCaHj527 foi submetido a
25 mutagênese sítio-dirigida por uma abordagem de PCR simples usando um sistema EXPAND® PCR de acordo com as instruções do fabricante. Nucleotídeos 134-144 foram convertidos de GTACTAAAACC (SEQ ID NO: 39) a CCGTTAAATTT (SEQ ID NO: 40) usando iniciador mutagênico 141223 mostrado abaixo.

Iniciador 141223:

5'-GGATGCTGTTGACTCCGGAAATTTAACGGTTTGGTCTTGCATC
CC-3' (SEQ ID NO: 41)

5 Nucleotídeos 423-436 foram convertidos de
ATGCAATTTAAACT (SEQ ID NO: 42) a CGGCAATTTAACGG (SEQ ID
NO: 43) usando iniciador mutagênico 141222 mostra abaixo.

Iniciador 141222:

5'-GGTATTGTCCTGCAGACGGCAATTTAACGGCTTCTGCGAATCGC-
3' (SEQ ID NO: 44)

10 O plasmídeo resultante foi designado pMT2188 (figura 3).

A região de codificação de endoglucanase V de *Humicola insolens* foi transferida de pCaHj170 como um fragmento *Bam* HI-*Sal* I em pMT2188 digerido com *Bam* HI e *Xho* I para gerar pCaHj568 (figura 4). Plasmídeo pCaHj568 compreende um promotor NA2-tpi mudado, ligado
15 operativamente à sequência de codificação de comprimento completo de endoglucanase V de *Humicola insolens*.

Exemplo 3: Construção de pMJ05

Plasmídeo pMJ05 foi construído por amplificação PCR da região de codificação de comprimento completo de endoglucanase V de
20 *Humicola insolens* de 915 bp de pCaHj568 usando iniciadores HiEGV-F e HiEGV-R mostrados abaixo.

HiEGV-F (sentido):

5'-AAGCTTAAGCATGCGTTCCTCCCCCTCC-3' (SEQ ID NO: 45)

HiEGV-R (anti-sentido)

25 5'-CTGCAGAATTCTACAGGCACTGATGGTACCAG-3' (SEQ ID NO: 46)

As reações de amplificação (50 µl) foram compostas de tampão de reação 1X ThermoPol (New England Biolabs, Beverly, MA, USA), 0,3 mM de dNTPs, 10 ng/µl de pCaHj568, 0,3 µM de iniciador HiEGV-F, 0,3 µM de iniciador HiEGV-R, e 2 unidades de Vent DNA

polimerase (New England Biolabs, Beverly, MA, USA). As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 5 ciclos cada durante 30 segundos a 94°C, 30 segundos a 50°C, e 60 segundos a 72°C, seguido por 25 ciclos cada durante 30 segundos a 94°C, 30 segundos a 65°C, e 120 segundos a 72°C (extensão final de 5 minutos). Os produtos de reação foram isolados em um gel de agarose a 1,0% usando tampão TAE onde uma banda de produto de 937 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

10 O fragmento 937 bp purificado foi usado como DNA gabarito para amplificações subsequentes usando os seguintes iniciadores:

HiEGV-R (anti-sentido):

5'-CTGCAGAATTCTACAGGCACTGATGGTACCAG-3' (SEQ ID NO: 47)

HiEGV-F-sobreposição (sentido):

15 5'-ACCGCGGACTGCGCATCATGCGTTCCTCCCCCTCC-3' (SEQ ID NO: 48)

Sequências do iniciador em itálico são homólogas para 17 bp do promotor do gene de celobiohidrolase I de *Trichoderma reesei* (*cbh1*) e sequências do iniciador sublinhadas são homólogos para 29 bp da região de codificação de endoglucanase V de *Humicola insolens*. Uma sobreposição de 36 bp entre o promotor e a sequência de codificação permitiu a fusão precisa do fragmento de 994 bp compreendendo o promotor de *Trichoderma reesei* *cbh1* para o fragmento de 918 bp compreendendo a região de codificação de endoglucanase V de *Humicola insolens*.

25 As reações de amplificação (50 µl) foram compostas de tampão de reação 1X ThermoPol, 0,3 mM de dNTPs, 1 µl do fragmento PCR de 937 bp purificado, 0,3 µM de iniciador de sobreposição HiEGV-F, 0,3 µM de iniciador HiEGV-R, e 2 unidades de Vent DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333

programado para 5 ciclos cada durante 30 segundos a 94°C, 30 segundos a 50°C, e 60 segundos a 72°C, seguido por 25 ciclos cada durante 30 segundos a 94°C, 30 segundos a 65°C, e 120 segundos a 72°C (extensão final de 5 minutos). Os produtos de reação foram isolados em um gel de agarose a 1,0% usando tampão TAE onde uma banda de produto de 945 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

Um PCR separado foi realizado para amplificar a sequência do promotor de *Trichoderma reesei cbh1* se estendendo de 994 bp a montante do códon de partida ATG do gene de DNA genômico RutC30 de *Trichoderma reesei* usando os iniciadores mostrados abaixo (o iniciador sentido foi engenheirado para ter um sítio de restrição *Sal* I na extremidade 5'). DNA genômico RutC30 de *Trichoderma reesei* foi isolado usando um kit DNEASY® Plant Maxi.

TrCBHIpro-F (sentido):

5'-AAACGTCGACCGAATGTAGGATTGTTATC-3' (SEQ ID NO: 49)

TrCBHIpro-R (anti-sentido):

5'-GATGCGCAGTCCGCGGT-3' (SEQ ID NO: 50)

As reações de amplificação (50 µl) foram compostas de tampão de reação 1X ThermoPol, 0,3 mM de dNTPs, 100 ng/µl de DNA genômico RutC30 de *Trichoderma reesei*, 0,3 µM de iniciador TrCBHIpro-F, 0,3 µM de iniciador TrCBHIpro-R, e 2 unidades de Vent DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada durante 30 segundos a 94°C, 30 segundos a 55°C, e 120 segundos a 72°C (extensão final de 5 minutos). Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 998 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

O fragmento PCR de 998 bp purificado foi usado como DNA gabarito para ampliações subsequentes usando os iniciadores mostrados abaixo.

TrCBH1pro-F:

5 5'-AAACGTCGACCGAATGTAGGATTGTTATC-3' (SEQ ID NO: 51)

TrCBH1pro-R-sobreposição:

5'-GGAGGGGGGAGGAACGCATGATGCGCAGTCCGCGGT-3' (SEQ ID NO: 52)

Sequências em itálico são homólogas para 17 bp do promotor *cbh1* de *Trichoderma reesei* e sequências sublinhadas são homólogas a 29 bp da região de codificação de endoglucanase V de *Humicola insolens*. Uma sobreposição de 36 bp entre o promotor e a sequência de codificação permitiu a fusão precisa do fragmento de 994 bp compreendendo o promotor *cbh1* de *Trichoderma reesei* para o fragmento de 918 bp compreendendo a região de codificação de comprimento completo de endoglucanase V de *Humicola insolens*.

As reações de amplificação (50 µl) foram compostas de tampão de reação 1X ThermoPol, 0,3 mM de dNTPs, 1 µl do fragmento PCR de 998 bp purificado, 0,3 µM TrCBH1pro-F iniciador, 0,3 µM iniciador de sobreposição TrCBH1pro-R, e 2 unidades de Vent DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 5 ciclos cada durante 30 segundos a 94°C, 30 segundos a 50°C, e 60 segundos a 72°C, seguido por 25 ciclos cada durante 30 segundos a 94°C, 30 segundos a 65°C, e 120 segundos a 72°C (extensão final de 5 minutos). Os produtos de reação foram isolados em um gel de agarose a 1,0% usando tampão TAE onde uma banda de produto de 1017 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

O fragmento PCR de promotor *cbh1* de *Trichoderma reesei* de

1017 bp e o fragmento PCR de endoglucanase V de *Humicola insolens* de 945 bp foram usados como DNA gabarito para amplificação subsequente usando os seguintes iniciadores para precisamente fusionar o promotor *cbh1* de 994 bp para a região de codificação de comprimento completo de endoglucanase V de 918 bp usando PCR de sobreposição.

TrCBH1pro-F:

5'-AAACGTCGACCGAATGTAGGATTGTTATC-3' (SEQ ID NO: 53)

HiEGV-R:

5'-CTGCAGAATTCTACAGGCACTGATGGTACCAG-3' (SEQ ID NO: 54)

10 As reações de amplificação (50 µl) foram compostas de tampão de reação 1X ThermoPol, 0,3 mM de dNTPs, 0,3 µM de iniciador TrCBH1pro-F, 0,3 µM de iniciador HiEGV-R, e 2 unidades de Vent DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 5 ciclos cada durante 30
15 segundos a 94°C, 30 segundos a 50°C, e 60 segundos a 72°C, seguido por 25 ciclos cada durante 30 segundos a 94°C, 30 segundos a 65°C, e 120 segundos a 72°C (extensão final de 5 minutos). Os produtos de reação foram isolados em um gel agarose a 1,0% usando Tampão TAE onde uma banda de produto de 1926 bp foi excisada do gel e purificada usando um kit de extração de gel
20 QIAQUICK® de acordo com as instruções do fabricante.

O fragmento de 1926 bp resultante foi clonado em um vetor pCR®-Blunt-II-TOPO® (Invitrogen, Carlsbad, CA, USA) usando um kit de clonagem PCR ZEROBLUNT® TOPO® (Invitrogen, Carlsbad, CA, USA) seguindo o protocolo de fabricação. O plasmídeo resultante foi digerido com
25 *Not* I e *Sal* I e o fragmento de 1926 bp foi purificado em gel usando um kit de extração de gel QIAQUICK® e ligado usando T4 DNA ligase (Roche, Indianápolis, IN, USA) em pMJ04, que também foi digerido com as mesmas duas enzimas de restrição, para gerar pMJ05 (figura 5). Plasmídeo pMJ05 compreende o promotor de celobiohidrolase I de *Trichoderma reesei* e

terminador ligado operativamente à sequência de codificação de comprimento completo de endoglucanase V de *Humicola insolens*.

Exemplo 4: Construção de vetor de expressão pSMai130

Um fragmento de DNA de 2586 bp se estendendo a partir do
 5 códon de partida ATG para o códon de parada TAA da sequência de
 codificação de comprimento completo de beta-glucosidase de *Aspergillus*
oryzae (SEQ ID NO: 21 para sequência de cDNA e SEQ ID NO:22 para a
 sequência de aminoácido deduzida; *E. coli* DSM 14240) foi amplificado por
 PCR de pJaL660 (WO 2002/095014) como gabarito com iniciadores 993467
 10 (sentido) e 993456 (anti-sentido) mostra abaixo. Um sítio *Spe* I foi
 engenheirado na extremidade 5' do iniciador anti-sentido para facilitar a
 ligação. Sequências do iniciador em itálico são homólogas para 24 bp do
 promotor de *cbh1* de *Trichoderma reesei* e sequências sublinhadas são
 homólogas a 22 bp à região de codificação de beta-glucosidase de *Aspergillus*
 15 *oryzae*.

Iniciador 993467:

5'-*ATAGTCAACCGCGGACTGCGCATCATGAAGCTTGGTTGGATCGA*
GG-3' (SEQ ID NO: 55)

Iniciador 993456:

20 5'-ACTAGTTTACTGGGCCTTAGGCAGCG-3' (SEQ ID NO: 56)

As reações de amplificação (50 µl) foram compostas de
 tampão de amplificação *Pfx* (Invitrogen, Carlsbad, CA, USA), 0,25 mM de
 dNTPs, 10 ng de pJaL660, 6,4 µM de iniciador 993467, 32 µM de iniciador
 993456, 1 mM de MgCl₂, e 2,5 unidades de *Pfx* DNA polimerase (Invitrogen,
 25 Carlsbad, CA, USA). As reações foram incubadas em um EPPENDORF®
 MASTERCYCLER® 5333 programado para 30 ciclos cada durante 1 minuto
 a 94°C, 1 minuto a 55°C, e 3 minutos a 72°C (extensão final de 15 minutos).
 Os produtos de reação foram isolados em um gel agarose a 1,0% usando
 tampão TAE onde uma banda de produto de 2586 bp foi excisada do gel e

purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

Um PCR separado foi realizado para amplificar a sequência de promotor de *cbh1* de *Trichoderma reesei* se estendendo de 1000 bp a montante do códon de partida ATG do gene, usando iniciador 993453 (sentido) e iniciador 993463 (anti-sentido) mostrados abaixo para gerar um fragmento PCR de 1000 bp.

Iniciador 993453:

5'-GTCGACTCGAAGCCCGAATGTAGGAT-3' (SEQ ID NO: 57) Iniciador 993463:

5'- CCTCGATCCAACCAAGCTTCATGATGCGCAGTCCGCGGTTG ACTA-3' (SEQ ID NO: 58)

Sequências do iniciador em itálico são homólogas para 24 bp do promotor de *cbh1* de *Trichoderma reesei* e sequências do iniciador sublinhadas são homólogas a 22 bp à região de codificação de comprimento completo de beta-glucosidase de *Aspergillus oryzae*. A sobreposição de 46 bp entre o promotor e a sequência de codificação permitiu a fusão precisa do fragmento de 1000 bp compreendendo o promotor de *cbh1* de *Trichoderma reesei* para o fragmento de 2586 bp compreendendo a região de codificação de beta-glucosidase de *Aspergillus oryzae*.

As reações de amplificação (50 µl) foram compostas de tampão de amplificação de *Pfx*, 0,25 mM de dNTPs, 100 ng de DNA genômico de RutC30 de *Trichoderma reesei*, 6,4 µM de iniciador 993453, 32 µM de iniciador 993463, 1 mM de MgCl₂, e 2,5 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada durante 1 minuto a 94°C, 1 minuto a 55°C, e 3 minutos a 72°C (extensão final de 15 minutos). Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 1000 bp foi excisada do gel e

purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

Os fragmentos purificados foram usados como DNA gabarito para amplificação subsequente por PCR de sobreposição usando iniciador 5 993453 (sentido) e iniciador 993456 (anti-sentido) mostrados acima para precisamente fundir o fragmento de 1000 bp compreendendo o promotor de *cbh1* de *Trichoderma reesei* para o fragmento de 2586 bp compreendendo a região de codificação de comprimento completo de beta-glucosidase de *Aspergillus oryzae*.

10 As reações de amplificação (50 µl) foram compostas de tampão de amplificação *Pfx*, 0,25 mM de dNTPs, 6,4 µM de iniciador 99353, 3,2 µM de iniciador 993456, 1 mM de MgCl₂, e 2,5 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada durante 1 minuto 15 a 94°C, 1 minuto a 60°C, e 4 minutos a 72°C (extensão final de 15 minutos).

O fragmento de 3586 bp resultante foi digerido com *Sal* I e *Spe* I e ligado em pMJ04, digerido com as mesmas duas enzimas de restrição, para gerar pSMai130 (figura 6). Plasmídeo pSMai130 compreende o promotor de gene de celobiohidrolase I de *Trichoderma reesei* e terminador 20 ligado operativamente à sequência sinal de beta-glucosidase nativa de *Aspergillus oryzae* e sequência de codificação (isto é, sequência de codificação de beta-glucosidase de *Aspergillus oryzae*).

Exemplo 5: Construção de pSMai135

25 A sequência de codificação de mature de beta-glucosidase de *Aspergillus oryzae* (menos a sequência de sinal nativa, ver figura 7; SEQ ID NOs: 59 e 60) de Lys-20 para o códon parado TAA foi amplificado por PCR a partir de pJaL660 como gabarito com iniciador 993728 (sentido) e iniciador 993727 (anti-sentido) mostra abaixo.

Iniciador 993728:

5'-TGCCGGTGTGGCCCTTGCCAAGGATGATCTCGCGTACTCCC-3'

(SEQ ID NO: 61)

Iniciador 993727:

5'-GACTAGTCTTACTGGGCCTTAGGCAGCG-3' (SEQ ID NO: 62)

5 Sequências em itálico são homólogas para 20 bp da sequência de sinal de endoglucanase V de *Humicola insolens* e sequências sublinhadas são homólogas a 22 bp à região de codificação de beta-glucosidase de *Aspergillus oryzae*. Um sítio *Spe* I foi engenheirado na extremidade 5' do iniciador anti-sentido.

10 As reações de amplificação (50 µl) foram compostas de tampão de amplificação *Pfx*, 0,25 mM de dNTPs, 10 ng/µl de pJaL660, 6,4 µM de iniciador 993728, 3,2 µM de iniciador 993727, 1 mM de MgCl₂, e 2,5 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada
15 durante 1 minuto a 94°C, 1 minuto a 55°C, e 3 minutos a 72°C (extensão final de 15 minutos). Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 2523 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

20 Uma amplificação PCR separada foi realizada para amplificar 1000 bp do promotor de *cbh1* de *Trichoderma reesei* e 63 bp à sequência de sinal de endoglucanase V de *Humicola insolens* (códon de partida ATG para Ala-21, figura 8, SEQ ID NOs: 63 e 64), usando iniciador 993724 (sentido) e iniciador 993729 (anti-sentido) mostrados abaixo.

25 Iniciador 993724:

5'-ACGCGTCGACCGAATGTAGGATTGTTATCC-3' (SEQ ID NO: 65)

Iniciador 993729:

5'-GGGAGTACGCGAGATCATCCTTGGCAAGGGCCAACACCGGCA-3'

(SEQ ID NO: 66)

Sequências do iniciador em itálico são homólogas para 20 bp à sequência de sinal de endoglucanase V de *Humicola insolens* e sequências de iniciador sublinhadas são homólogas para 22 bp à região de codificação de beta-glucosidase de *Aspergillus oryzae*.

5 Plasmídeo pMJ05, que compreende a região de codificação de endoglucanase V de *Humicola insolens* sob o controle do promotor *cbh1*, foi usado como gabarito para gerar um fragmento de 1063 bp compreendendo o promotor de *cbh1* de *Trichoderma reesei* e fragmento de sequência de sinal de endoglucanase V de *Humicola insolens*. Uma sobreposição de 42 bp foi
10 compartilhada entre o promotor *cbh1* de *Trichoderma reesei* e a sequência de sinal de endoglucanase V de *Humicola insolens* e a sequência de codificação mature de beta-glucosidase de *Aspergillus oryzae* para prover uma ligação perfeita entre o promotor e o códon de partida ATG à região de codificação de beta-glucosidase de 2523 bp de *Aspergillus oryzae*.

15 As reações de amplificação (50 µl) foram compostas de tampão de amplificação *Pfx*, 0,25 mM de dNTPs, 10 ng/µl de pMJ05, 6,4 µM de iniciador 993728, 32 µM de iniciador 993727, 1 mM de MgCl₂, e 2,5 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada
20 durante 1 minuto a 94°C, 1 minuto a 60°C, e 4 minutos a 72°C (extensão final de 15 minutos). Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 1063 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

25 Os fragmentos de sobreposição purificados foram usados como gabaritos para amplificação usando iniciador 993724 (sentido) e iniciador 993727 (anti-sentido) descritos acima para precisamente fusionar o fragmento de 1063 bp compreendendo o promotor *cbh1* de *Trichoderma reesei* e sequência de sinal de endoglucanase V de *Humicola insolens* para o

fragmento de 2523 bp compreendendo a armação da região de codificação madura de beta-glucosidase de *Aspergillus oryzae* por PCR de sobreposição.

As reações de amplificação (50 µl) foram compostas de tampão de amplificação *Pfx*, 0,25 mM de dNTPs, 6,4 µM de iniciador 993724, 3,2 µM de iniciador 993727, 1 mM de MgCl₂, e 2,5 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada durante 1 minuto a 94°C, 1 minuto a 60°C, e 4 minutos a 72°C (extensão final de 15 minutos). Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 3591 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

O fragmento de 3591 bp resultante foi digerido com *Sal* I e *Spe* I e ligado em pMJ04 digerido com as mesmas enzimas de restrição para gerar pSMai135 (figura 9). Plasmídeo pSMai135 compreende o promotor de gene de celobiohidrolase I de *Trichoderma reesei* e terminador ligado operativamente à sequência de sinal de endoglucanase V de *Humicola insolens* e a sequência de codificação madura de beta-glucosidase de *Aspergillus oryzae*.

20 **Exemplo 6: Expressão de beta-glucosidase de *Aspergillus oryzae* com o sinal de secreção de endoglucanase V de *Humicola insolens***

Plasmídeo pSMai135 codificando a beta-glucosidase madura de *Aspergillus oryzae* ligada ao sinal de secreção de endoglucanase V de *Humicola insolens* (figura 8) foi introduzido em RutC30 de *Trichoderma reesei* por transformação mediada de PEG (Penttila *et al.*, 1987, *Gene* 61 155 - 164). O plasmídeo continha o gene de *amdS* de *Aspergillus nidulans* para permitir aos transformantes crescer em acetamida como a única fonte de nitrogênio.

RutC30 de *Trichoderma reesei* foi cultivado a 27°C e 90 rpm

em 25 ml de meio YP suplementado com 2% (p/v) glicose e 10 mM de uridina durante 17 horas. Micélios foram coletados por filtração usando um sistema de filtração descartável acionado a vácuo (Millipore, Bedford, MA, USA) e lavado duas vezes com água deionizada e duas vezes com 1,2 M de sorbitol. Protoplastos foram gerados por colocação em suspensão do micélio lavado em 20 ml de 1,2 M de sorbitol contendo 15 mg de GLUCANEX® (Novozymes A/S, Bagsvaerd, Dinamarca) por ml e 03,6 unidades de quitinase (Sigma Chemical Co., St. Louis, MO, USA) por ml e incubando durante 15-25 minutos a 34°C com agitação suave a 90 rpm. Protoplastos foram coletados por centrifugar durante 7 minutos a 400 x g e lavados duas vezes com 1,2 M de sorbitol frio. Os protoplastos foram contados usando um hemacitômetro e re-colocados em suspensão em STC para uma concentração final de 1×10^8 protoplastos por ml. Os protoplastos em excesso foram armazenados em um recipiente de congelamento Cryo 1°C (Nalgene, Rochester, NY, USA) a -80°C.

Aproximadamente 7 µg de pSMai135 digerido com *Pme* I foram adicionados a 100 µl de solução de protoplasto e misturado suavemente, seguido por 260 µl de tampão PEG, misturado, e incubados em temperatura ambiente durante 30 minutos. STC (3 ml) foi então adicionado e misturado e a solução de transformação foi colocada sobre placas COVE usando seleção de *amsdS* de *Aspergillus nidulans*. As placas foram incubadas a 28°C durante 5-7 dias. Transformantes foram sub-cultivados sobre placas COVE2 e cultivados a 28°C.

Sessenta e sete transformantes designados SMA135 obtidos com pSMai135 foram subcultivados sobre placas congeladas contendo acetamida e deixados esporular durante 7 dias a 28°C.

Sessenta e sete transformantes de SMA135 de *Trichoderma reesei* foram cultivados em 125 ml de frascos com agitação com anteparos contendo 25 ml de meio induzindo celulase a pH 6,0 inoculado com esporos

dos transformantes e incubados a 28°C e 200 rpm durante 7 dias. RutC30 de *Trichoderma reesei* foi ciclado como um controle. Amostras do caldo de cultura foram removidas no dia 7. Um ml de cada caldo de cultura foi centrifugado a 15.700 x g durante 5 minutos em uma micro-centrífuga e os sobrenadantes transferidos para novos tubos. As amostras foram armazenadas a 4°C até teste de enzima. Os sobrenadantes foram testados para atividade de beta-glucosidase usando p-nitrofenil-beta-D-glucopiranosídeo como substrato, como descrito abaixo.

Atividade de beta-glucosidase foi determinada em temperatura ambiente usando 25 µl de alíquotas de sobrenadantes de cultura, diluída 1:10 em 50 mM de succinato pH 5,0, em 200 µl de 0,5 mg/ml de p-nitrofenil-beta-D-glucopiranosídeo como substrato em 50 mM de succinato pH 5,0. Após 15 minutos de incubação a reação foi parada por adição de 100 µl de 1 M de Tris-HCl pH 8,0 e a absorção foi lida espectrofotometricamente a 405 nm. Uma unidade de atividade de beta-glucosidase correspondendo à produção de 1 µmol de p-nitrofenila por minuto por litro a pH 5,0, temperatura ambiente. Beta-glucosidase de *Aspergillus niger* (NOVOZYM™ 188, Novozymes A/S, Bagsvaerd, Dinamarca) foi usado como um padrão de enzima.

Um número dos transformantes de SMA135 mostrou atividades de beta-glucosidase várias vezes maior do que a secretada por RutC30 de *Trichoderma reesei*. Dentre os transformantes SMA135 triados, transformante SMA135-04 produziu a maior atividade de beta-glucosidase.

SDS-PAGE foi realizado usando Tris-HCl CRITERION® (5% resolução) géis (Bio-Rad, Hercules, CA, USA) com o sistema CRITERION® (Bio-Rad, Hercules, CA, USA). Cinco µl de sobrenadantes de dia 7 (ver acima) foram colocados em suspensão em 2X concentração de tampão de amostra Laemmli (Bio-Rad, Hercules, CA, USA) e fervidos na presença de 5% beta-mercaptoetanol durante 3 minutos. As amostras de sobrenadante foram carregadas sobre um gel de poliacrilamida e submetidas a eletroforese

com 1X de Tris/glicina/SDS como tampão de ciclo (Bio-Rad, Hercules, CA, USA). O gel resultante foi colorido com corante Coomassie BIO-SAFE® (Bio-Rad, Hercules, CA, USA).

5 Dentre os transformantes de SMA13538 *Trichoderma reesei* analisados por SDS-PAGE, 26 produziram uma proteína de aproximadamente 110 kDa que não foi visível em RutC30 de *Trichoderma reesei* como controle. Transformante *Trichoderma reesei* SMA135-04 produziu o maior nível de beta-glucosidase como evidenciado por abundância da banda de 110 kDa vista por SDS-PAGE.

10 **Exemplo 7: Construção de vetor de expressão pSMai140**

Vetor de expressão pSMai140 foi construído por digestão de plasmídeo pSATE111BG41 (WO 04/099228), que transporta uma região de codificação de comprimento completo de variante BG41 de beta-glucosidase de *Aspergillus oryzae* (SEQ ID NO: 23 que codifica a sequência de aminoácido de SEQ ID NO: 24), com *Nco* I. O fragmento de 1243 bp resultante foi isolado em um gel agarose a 1,0% usando tampão TAE e purificado usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

20 O vetor de expressão pSMai135 foi digerido com *Nco* I e um fragmento de 8286 bp foi isolado em um gel agarose a 1,0% usando tampão TAE e purificado usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante. O fragmento de variante BG41 de beta-glucosidase de 1243 bp *Nco* I de *Aspergillus oryzae* digerido foi então ligado no vetor 8286 bp, usando T4 DNA ligase (Roche, Indianápolis, IN, USA) de acordo com o protocolo de fabricação, para criar o vetor de expressão pSMai140 (figura 10). Plasmídeo pSMai140 compreende o promotor de gene de celobiohidrolase I (CEL7A) de *Trichoderma reesei* e terminador ligado operativamente na sequência de sinal de endoglucanase V de *Humicola insolens* e a sequência de codificação mature variante de beta-glucosidase de

Aspergillus oryzae.

Exemplo 8: Transformação de RutC30 de *Trichoderma reesei* com pSMai140

5 Plasmídeo pSMai140 foi linearizado com *Pme* I e transformado na cepa de RutC30 de *Trichoderma reesei* como descrito no exemplo 6. Um total de 100 transformantes foram obtidos a partir de quatro experimentos de transformação independentes, todos sendo cultivados em frascos de agitação em meio induzindo celulase, e a atividade de beta-glucosidase foi medida a partir do meio de cultura dos transformantes como
10 descrito no exemplo 6. Um número de transformantes de SMA140 de *Trichoderma reesei* mostrou atividades de beta-glucosidase várias vezes maiores do que a de RutC30 de *Trichoderma reesei*.

A presença da proteína de variante BG41 de beta-glucosidase de *Aspergillus oryzae* no meio de cultura foi detectada por eletroforese de gel
15 de SDS-poliacrilamida como descrito no exemplo 6 e a coloração Coomassie dos mesmos 13 sobrenadantes de cultura dos quais as atividades de enzimas foram analisadas. Todos os treze transformantes que têm atividade de β -glucosidase elevada, também expressaram a variante BG41 de aproximadamente 110 KDa de beta-glucosidase de *Aspergillus oryzae*, em
20 rendimentos variados.

O transformante expressando variante de beta-glucosidase maior, como avaliado por teste de atividade de beta-glucosidase e eletroforese gel de SDS-poliacrilamida, foi designado SMA140-43 *Trichoderma reesei*.

Exemplo 9: Construção de vetor de expressão pSaMe-F1

25 Um fragmento de DNA contendo 209 bp do promotor de gene de celobiohidrolase I de *Trichoderma reesei* e a região do núcleo (nucleotídeos 1 a 702 de SEQ ID NO: 1 que codifica aminoácidos 1 a 234 de SEQ ID NO: 2; WO 91/17243) do gene de endoglucanase V de *Humicola insolens* foi amplificado por PCR usando pMJ05 como gabarito usando os

iniciadores mostrados abaixo.

995103:

5'-cccaagcttagccaagaaca-3' (SEQ ID NO: 67)

995137:

5 5'-gggggaggaacgcatgggatctggacggc-3' (SEQ ID NO: 68)

As reações de amplificação (50 µl) foram compostas de tampão de amplificação 1X *Pfx*, 10 mM de dNTPs, 50 mM de MgSO₄, 10 ng/µl de pMJ05, 50 picomoles de iniciador 995103, 50 picomoles de iniciador 995137, e 2 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada durante 30 segundos a 94°C, 30 segundos a 55°C, e 60 segundos a 72°C (extensão final de 3 minutos).

Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 911 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

Um fragmento de DNA contendo 806 bp do gene de variante BG41 de beta-glucosidase de *Aspergillus oryzae* foi PCR amplificado usando pSMai140 como gabarito e os iniciadores mostrados abaixo.

20 995133:

5'-gccgtccagatccccatgcttctctcccc-3' (SEQ ID NO: 69)

995111:

5'-ccaagcttggtcagagtttc-3' (SEQ ID NO: 70)

As reações de amplificação (50 µl) foram compostas de tampão de amplificação 1X *Pfx*, 10 mM de dNTPs, 50 mM MgSO₄, 100 ng de pSMai140, 50 picomoles de iniciador 995133, 50 picomoles de iniciador 995111, e 2 unidades de *Pfx* DNA polimerase. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para 30 ciclos cada durante 30 segundos a 94°C, 30 segundos a 55°C, e 120 segundos

a 72°C (extensão final de 3 minutos).

Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 806 bp foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

Os dois fragmentos PCR acima foram então submetidos a PCR de sobreposição. Os fragmentos de sobreposição purificados foram usados como gabaritos para amplificação usando iniciador 995103 (sentido) e iniciador 995111 (anti-sentido) descritos acima para precisamente fusionar o fragmento de 702 bp compreendendo 209 bp do promotor de gene de celobiohidrolase I de *Trichoderma reesei* e a sequência de núcleo de endoglucanase V de *Humicola insolens* para o fragmento de 806 bp compreendendo uma porção da região de codificação de variante BG41 de beta-glucosidase de *Aspergillus oryzae* por PCR de sobreposição.

As reações de amplificação (50 µl) foram compostas de tampão de amplificação 1X *Pfx*, 10 mM de dNTPs, 50 mM de MgSO₄, 2,5 µl de cada fragmento (20 ng/µl), 50 picomoles de iniciador 995103, 50 picomoles de iniciador 995111, e 2 unidades de *Pfx* DNA polimerase de alta fidelidade. As reações foram incubadas em um EPPENDORF® MASTERCYCLER® 5333 programado para uma desnaturação inicial de 3 minutos a 95°C seguido por 30 ciclos cada durante 1 minuto de desnaturação, 1 minuto anelando a 60°C, e uma extensão de 3 minutos a 72°C.

Os produtos de reação foram isolados em um gel agarose a 1,0% usando tampão TAE onde uma banda de produto de 1,7 kb foi excisada do gel e purificada usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante.

O fragmento de 1,7 kb foi ligado em um vetor de pCR®4 Blunt (Invitrogen, Carlsbad, CA, USA) de acordo com as instruções do fabricante. A construção foi então transformada em células *E. coli* ONE

SHOT® TOP10 quimicamente competentes (Invitrogen, Carlsbad, CA, USA) de acordo com o procedimento de transformação de produtos químicos rápido do fabricante. Colônias foram selecionadas e analisadas por isolamento de plasmídeo e digestão com *Hind* III para liberar o fragmento PCR de sobreposição de 1,7 kb.

Plasmídeo pSMai140 também foi digerido com *Hind* III para linearizar o plasmídeo. Ambos fragmentos digeridos foram combinados em uma reação de ligação usando um kit de ligação de DNA rápido seguindo as instruções do fabricante para produzir pSaMe-F1 (figura 11).

Células competentes de *E. coli* XL1-Blue Subcloning-Grade (Stratagene, La Jolla, CA, USA) foram transformadas com o produto de ligação. A identidade da construção foi confirmada por Sequenciamento de DNA do promotor de gene de celobiohidrolase I de *Trichoderma reesei*, sequência de sinal de endoglucanase V de *Humicola insolens*, núcleo de endoglucanase V de *Humicola insolens*, sequência de sinal de endoglucanase V de *Humicola insolens*, variante de BG41 de beta-glucosidase de *Aspergillus oryzae*, e a sequência de terminador de gene de celobiohidrolase I de *Trichoderma reesei* de plasmídeos purificados de *E. coli* transformados. Um clone contendo o plasmídeo recombinante foi designado pSaMe-F1.

Plasmídeo pSaMe-F1 compreende o promotor de gene de celobiohidrolase I de *Trichoderma reesei* e terminador e a sequência de peptídeo de sinal de endoglucanase V de *Humicola insolens* ligado diretamente no polipeptídeo de núcleo de endoglucanase V de *Humicola insolens* que são diretamente fusionados no peptídeo de sinal de endoglucanase V de *Humicola insolens* que é diretamente ligado na sequência de codificação madura de variante BG41 de beta-glucosidase de *Aspergillus oryzae*. A sequência de DNA e a sequência de aminoácido deduzidas da proteína de fusão de variante BG de beta-glucosidase de *Aspergillus oryzae* são mostradas em SEQ ID NOs: 73 e 74, respectivamente (ver figuras 14A, 14B, 14C, e 14D).

Exemplo 10: Transformação de RutC30 de *Trichoderma reesei* com pSaMe-F1

5 Frascos de agitação contendo 25 ml de meio YP suplementado com 2% de glicose e 10 mM de uridina foram inoculados com 5×10^7 esporos de RutC30 de *Trichoderma reesei*. Seguindo incubação durante a noite durante aproximadamente 16 horas a 27°C, 90 rpm, os micélios foram coletados usando um sistema de filtração descartável acionado por vácuo. Os micélios foram lavados duas vezes em 100 ml de água deionizada e duas vezes em 1,2 M de sorbitol. Protoplastos foram gerados como descrito em exemplo 6.

10 Duas microgramas de pSaMe-F1 DNA linearizado com *Pme* 1, 100 µl de protoplastos de RutC30 de *Trichoderma reesei*, e 50% de PEG (4000) foram misturados e incubados durante 30 minutos em temperatura ambiente. Então 3 ml de STC foram adicionados e os conteúdos foram despejados sobre uma placa COVE suplementada com 10 mM de uridina. A placa foi então incubada a 28°C. Transformantes começaram a aparecer pelo dia 6 e foram tirados das placas COVE2 durante crescimento a 28°C e 6 dias. Vinte e dois transformantes de *Trichoderma reesei* foram recuperados.

20 Transformantes foram cultivados em frascos de agitação em meio induzindo celulase, e atividade de beta-glucosidase foi medida como descrito no exemplo 6. Um número de transformantes pSaMe-F1 produziu atividade de beta-glucosidase. Um transformante, designado de *Trichoderma reesei* SaMeF1-9, produziu a quantidade maior de beta- glucosidase, e tem duas vezes a atividade de uma cepa expressando a variante de beta-glucosidase de *Aspergillus oryzae* (exemplo 9).

25 A atividade de endoglucanase foi testada usando um teste de deposição de carboximetil celulose (CMC), de acordo com Beguin, 1983, *Analytical Biochem.* 131(2): 333 - 336. Cinco µg de proteína total dentre as cinco amostras do caldo (as tendo a atividade de beta-glucosidase maior)

foram diluídos em tampão de amostra nativo (Bio-Rad, Hercules, CA, USA) e ciclados em um gel Tris-HCl CRITERION® 8-16% (Bio-Rad, Hercules, CA, USA) usando 10X de tampão de ciclo Tris/glicina (Bio-Rad, Hercules, CA, USA) e então o gel foi depositado no topo de uma placa contendo 1% de carboximetilcelulose (CMC). Após 1 hora de incubação a 37°C, o gel foi colorido com 0,1% Congo Red durante 20 minutos. A placa foi então descolorida usando 1 M de NaCl a fim de identificar regiões de depuração indicativas de atividade de endoglucanase. Duas zonas de depuração foram visíveis, uma zona superior a aproximadamente 110 kDa e uma zona inferior a aproximadamente 25 kDa. O tamanho de proteína previsto da fusão de variante BG41 de endoglucanase V de *Humicola insolens* e beta-glucosidase de *Aspergillus oryzae* é de 118 kDa se as duas proteínas não forem clivadas e permanecerem como um polipeptídeo único, glicosilação do domínio do núcleo de endoglucanase V individual e dos fios de beta-glucosidase para migração das proteínas individuais em um maior peso molecular do que previsto a partir da sequência primária. Se as duas proteínas são clivadas então os tamanhos previstos para o domínio de núcleo de endoglucanase V de *Humicola insolens* é de 24 kDa e 94 kDa para variante BG41 de beta-glucosidase de *Aspergillus oryzae*. Porque se tem uma zona de depuração em aproximadamente 110 kDa, este resultado indicou que minimamente uma população da proteína de fusão de endoglucanase e beta-glucosidase permanece intacta como uma proteína grande única. A zona de depuração inferior representa, o mais provavelmente, uma atividade de endoglucanase endógena, e possivelmente adicionalmente resulta de clivagem parcial do domínio do núcleo de endoglucanase V de *Humicola insolens* a partir de β -glucosidase de *Aspergillus oryzae*.

Os resultados demonstraram que o núcleo de endoglucanase V de *Humicola insolens* estava ativo mesmo enquanto fusionado a beta-glucosidase de *Aspergillus oryzae*. Além disso, o aumento em atividade de

beta-glicosidase pareceu resultar de secreção aumentada de proteína com relação à eficiência de secreção da beta-glicosidase não fusão. Por ligação da sequência de variante BG41 de beta-glicosidase de *Aspergillus oryzae* para o núcleo de endoglucanase V de *Humicola insolens* eficientemente secretado, mais beta-glicosidase foi secretada.

Exemplo 11: Construção de vetor pSaMe-FX

Plasmídeo pSaMe-FX foi construído por modificação de pSaMe-F1. Plasmídeo pSaMe-F1 foi digerido com *Bst* Z17 e *Eco* RI para gerar um fragmento de 1 kb que continha a sequência de codificação de variante BG41 de beta-glicosidase e um fragmento de 9,2 kb contendo o restante do plasmídeo. Os fragmentos foram separados em um gel agarose a 1,0% usando tampão TAE e o fragmento de 9,2 kb foi excisado e purificado usando um kit de extração de gel QIAQUICK® de acordo com as instruções do fabricante. Plasmídeo pSMai135 também foi digerido com *Bst* Z17 e *Eco* RI para gerar um fragmento de 1 kb contendo bases homólogas à sequência de codificação de variante BG41 de beta-glicosidase de *Aspergillus oryzae* e um fragmento de 8,5 kb contendo o restante do plasmídeo. O fragmento de 1 kb foi isolado e purificado como acima.

Os fragmentos de 9,2 kb e 1 kb foram combinados em uma reação de ligação usando um kit de ligação de DNA rápido seguindo as instruções do fabricante para produzir pSaMe-FX, que é idêntico para pSaMe-F1 exceto que continha a sequência de codificação madura de beta-glicosidase de tipo selvagem em vez de uma sequência de codificação madura variante.

Células competentes de *E. coli* SURE® (Stratagene, La Jolla, CA, USA) foram transformadas com o produto de ligação. A identidade da construção foi confirmada por sequenciamento de DNA do promotor de gene de celobiohidrolase I de *Trichoderma reesei*, sequência de sinal de endoglucanase V de *Humicola insolens*, sequência de núcleo de

endoglucanase V de *Humicola insolens*, sequência de sinal de endoglucanase V de *Humicola insolens*, sequência de codificação madura de beta-glucosidase de *Aspergillus oryzae*, e a sequência de terminador de gene de celobiohidrolase I de *Trichoderma reesei* de plasmídeos purificados a partir de *E. coli* transformado. Um clone contendo o plasmídeo recombinante foi designado pSaMe-FX (figura 12). A sequência de DNA e a sequência de aminoácido deduzidas da proteína de fusão de beta-glucosidase de *Aspergillus oryzae* são mostradas nas SEQ ID NOs: 75 e 76, respectivamente (ver figuras 15A, 15B, 15C, e 15D).

10 **Exemplo 12: Transformação e expressão de transformantes *Trichoderma***

A construção de pSaMe-FX foi linearizada com *Pme* I e transformada na cepa RutC30 de *Trichoderma reesei* como descrito no exemplo 10. Um total de 63 transformantes foi obtido a partir de uma transformação única. Os transformantes foram cultivados em frascos de agitação em meio induzindo celulase, e atividade de beta-glucosidase foi medida como descrito no exemplo 6. Um número de transformantes pSaMe-FX produziu atividade de beta-glucosidase. Um transformante designado SaMe-FX16 produziu duas vezes a quantidade de atividade de beta-glucosidase comparado a SaMeF1-9 de *Trichoderma reesei* (exemplo 10).

20 **Exemplo 13: Análise de transformantes de *Trichoderma reesei***

Uma proteína de fusão foi construída como descrito no exemplo 9 por fusão do núcleo de endoglucanase V de *Humicola insolens* (contendo sua sequência de sinal própria nativa) com a sequência de codificação madura de variante BG41 de beta-glucosidase de *Aspergillus oryzae* ligada à sequência de sinal de endoglucanase V de *Humicola insolens*. Esta construção de fusão resultou em um aumento de duas vezes em atividade de beta-glucosidase secretada comparado à sequência de codificação madura de variante BG41 de beta-glucosidase de *Aspergillus oryzae* ligada à sequência de sinal de endoglucanase V de *Humicola insolens*. Uma segunda

construção de fusão foi feita como descrito no exemplo 11 consistindo do núcleo de endoglucanase V de *Humicola insolens* (contendo sua sequência de sinal própria) fusionado com a sequência de codificação de beta-glucosidase de tipo selvagem de *Aspergillus oryzae* ligada à sequência de sinal de endoglucanase V de *Humicola insolens*, e isto levou a uma melhora ainda maior na atividade de beta-glucosidase. A cepa transformada com a fusão de tipo selvagem tem duas vezes a atividade de beta-glucosidase secretada relativa à cepa transformada com a fusão de variante BG41 de beta-glucosidase.

10 **Exemplo 14: Clonagem da sequência de codificação de proteína de fusão de beta-glucosidase em um vetor de expressão de *Aspergillus oryzae***

Dois iniciadores de oligonucleotídeo sintéticos, mostrados abaixo, foram designados para amplificar PCR a matriz de leitura aberta de comprimento completo de pSaMeFX codificando a proteína de fusão de beta-glucosidase.

Iniciador dianteiro PCR:

5'-GGACTGCGCAGCATGCGTTC-3' (SEQ ID NO: 71)

Iniciador reverso PCR:

5'-AGTTAATTAATTACTGGGCCTTAGGCAGCG-3' (SEQ ID NO: 72)

20 As letras em negrito representam a sequência de codificação. O "G" sublinhado no iniciador dianteiro representa uma mudança de base introduzida para criar o sítio de restrição *Sph* I. A sequência restante contém a identidade de sequência comparada com os sítios de inserção de pSaMeFX. A sequência sublinhada no iniciador reverso representa um sítio de restrição *Pac* I adicionado para facilitar a clonagem deste gene no vetor de expressão pAILo2 (WO 04/099228).

Cinquenta picomoles de cada dos iniciadores acima foram usados em uma reação de PCR contendo 50 ng de pSaMeFX DNA, 1X de tampão de amplificação *Pfx*, 6 µl de 10 mM de mistura de dATP, dTTP,

dGTP, e dCTP, 2,5 unidades de PLATINUM® *Pfx* DNA Polimerase, e 1 µl de 50 mM de MgSO₄ em um volume final de 50 µl. Um EPPENDORF® MASTERCYCLER® 5333 foi usado para amplificar o fragmento programado durante 1 ciclo a 98°C durante 2 minutos; e 35 ciclos cada a 5 96°C durante 30 segundos, 61°C durante 30 segundos, e 68°C durante 3 minutos. Após os 35 ciclos, a reação foi incubada a 68°C durante 10 minutos e então resfriada a 10°C até ser ainda processado. Um produto de reação de PCR de 3,3 kb foi isolado em um GTG-agarose a 0,8% (Cambrex Bioproducts One Meadowlands Plaza East Rutherford, NJ, USA) usando 10 tampão TAE e 0,1 µg de brometo de etídio por ml. O DNA foi visualizado com o auxílio de um DARK READER™ (Clare Chemical Research, Dolores, CO, USA) para evitar mutações induzidas por UV. Uma banda de DNA de 3,3 kb foi excisada com uma lâmina de barbear descartável e purificada com um copo de girar ULTRAFREE®-DA (Millipore, Billerica, 15 MA, USA) de acordo com as instruções do fabricante.

O produto PCR de 3,3 kb purificado foi clonado em um vetor de pCR®4Blunt-TOPO® (Invitrogen, Carlsbad, CA, USA). Quatro microlitros do produto PCR purificado foram misturados com 1 µl de uma solução de 2 M de cloreto de sódio e 1 µl do vetor TOPO®. A reação foi 20 incubada em temperatura ambiente durante 15 minutos e então 2 µl da reação foram usados para transformar células de *E. coli* One Shot® TOP10 quimicamente competentes, de acordo com as instruções do fabricante. Três alíquotas de 83 µl cada da reação de transformação foram espalhadas sobre as três placas de 150 mm 2X YT suplementadas com 100 µg de ampicilina por 25 ml e incubadas durante a noite a 37°C.

Oito colônias recombinantes foram usadas para inocular culturas líquidas contendo 3 ml de meio LB suplementado com 100 µg de ampicilina por ml. DNA de plasmídeo foi preparado a partir destas culturas usando um BIOROBOT® 9600 (QIAGEN Inc., Valencia, CA, USA). Clones

foram analisados por digestão de enzima de restrição com *Pac* I. DNA de plasmídeo de cada clone foi digerido com *Pac* I e analisado por eletroforese de gel agarose a 1,0% usando tampão TAE. Todos os oito clones têm o padrão de digestão de restrição esperado e clones 5, 6, 7, e 8 foram selecionados para serem sequenciados para confirmar que não se tem mutações no inserto clonado. Análise de sequência de suas extremidades 5' e 3' indicam que todos os 4 clones têm a sequência correta. Os clones 5 e 7 foram selecionados para outro sequenciamento. Ambos os clones foram sequenciados para valores Phred Q maiores do que 40 para assegurar que não se tem erros induzidos por PCR. Os clones 5 e 7 foram mostrados como tendo a sequência esperada e o clone 5 foi selecionado para re-clonagem em pAILo2.

DNA de plasmídeo de clone 5 foi linearizado por digestão com *Sph* I. O clone linearizado foi então terminado de modo cego por adição de 1,2 µl de uma mistura a 10 mM de dATP, dTTP, dGTP, e dCTP e 6 unidades de T4 DNA polimerase (New England Biolabs, Inc., Ipswich, MA, USA). A mistura foi incubada a 12°C durante 20 minutos e então a reação foi parada por adição de 1 µl de 0,5 M de EDTA e aquecimento a 75°C durante 20 minutos para inativar a enzima. Um fragmento de 3,3 kb codificando a proteína de fusão de beta-glucosidase foi purificado por eletroforese gel e ultrafiltração como descrito acima.

O vetor pAILo2 foi linearizado por digestão com *Nco* I. O vetor linearizado foi então terminado cego por adição de 0,5 µl de uma mistura a 10 mM de dATP, dTTP, dGTP, e dCTP e uma unidade de DNA polimerase I. A mistura foi incubada a 25°C durante 15 minutos e então a reação foi parada por adição de 1 µl de 0,5M de EDTA e aquecimento a 75°C durante 15 minutos para inativar as enzimas. Então o vetor foi digerido com *Pac* I. O vetor terminado de modo cego foi purificado por gel de eletroforese e ultrafiltração como descrito acima.

Clonagem do fragmento de 3,3 kb codificando a proteína de fusão de beta-glucosidase no vetor pAIIo2 linearizado e purificado foi realizada com um kit de ligação rápido. Uma amostra de 1 µl da reação foi usada para transformar as células de *E. coli* XL10 SOLOPACK® Gold (Stratagene, La Jolla, CA, USA) de acordo com as instruções do fabricante. Após o período recuperado, duas alíquotas de 100 µl de uma reação de transformação foram colocadas sobre duas placas de 150 mm 2X YT suplementadas com 100 µg de ampicilina por ml e incubadas durante a noite a 37°C. Um conjunto de oito clones recombinantes putativos foi selecionado de modo aleatório dentre as placas de seleção e DNA de plasmídeo foi preparado dentre cada um usando um BIOROBOT® 9600. Os clones 1-4 foram selecionados para sequenciamento com iniciadores pAIIo2 específicos para confirmar que a junção vetor/inserto apresenta a sequência correta. O clone 3 tem uma junção de vetor/inserto perfeita e foi designado pAIIo47 (figura 13).

A fim de criar uma cepa de expressão isenta de marcador, uma digestão de endonuclease de restrição foi realizada para separar o gene *blaA* que confere resistência ao antibiótico ampicilina dentre o resto da construção de expressão. Trinta microgramas de pAIIo47 foram digeridos com *Pme* I. O DNA digerido foi então purificado por eletroforese gel agarose como descrito acima. Uma banda de DNA de 6,4 kb contendo a construção de expressão mas faltando o gene *blaA* foi excisada com uma lâmina de barbear e purificada com um kit de extração de gel QIAQUICK®.

Exemplo 15: Expressão da proteína de fusão de beta-glucosidase em JaL355 de *Aspergillus oryzae*

Protoplastos de JaL355 de *Aspergillus oryzae* (WO 00/240694) foram preparados de acordo com o método de Christensen *et al.*, 1988, *Bio/Technology* 6: 1419 - 1422. Dez microlitros da construção de expressão purificado de exemplo 14 foram usados para transformar protoplastos de JaL355 de *Aspergillus oryzae*. A transformação de JaL355 de

Aspergillus oryzae rendeu aproximadamente 90 transformantes. Cinquenta transformantes foram isolados para placas de PDA individuais e incubados durante cinco dias a 34°C.

5 Quarenta e oito placas de esporos confluentes foram lavadas com 3 ml de 0,01% de TWEEN® 80 e a suspensão de esporo foi usada para inocular 25 ml de meio MDU2BP em 125 ml de frascos de agitação de vidro. Culturas transformantes foram incubadas a 34°C com agitação constante a 200 rpm. Após 5 dias, 1 ml de alíquotas de cada cultura foi centrifugado a 12.000 x g e seus sobrenadantes coletados. Cinco µl de cada sobrenadante
10 foram misturados com um volume igual de tampão de carga 2X (10% de beta-mercaptoetanol) e carregados sobre um gel SDS-PAGE 1,5 mm 8%-16 % Tris-glicina e coloridos com coloração de proteína tipo BIO-SAFE® Coomassie Blue G250 (Bio-Rad, Hercules, CA, USA). Os perfis de SDS-PAGE dos caldos de cultura mostraram que 33 dentre os 48 transformantes
15 foram capazes de expressar uma proteína nova com um peso molecular aparente muito próximo do esperado de 118 kDa. Transformante 21 produziu o melhor rendimento e foi selecionado para outros estudos.

Exemplo 16: Isolamento de esporo único de transformante 21 de JaL355 de *Aspergillus oryzae*

20 Os esporos do transformante 21 de JaL355 de *Aspergillus oryzae* foram espalhados sobre uma placa de PDA e incubados durante cinco dias a 34°C. Uma pequena área da placa de esporo confluyente foi lavada com 0,5 ml de 0,01% de TWEEN® 80 para recolocar em suspensão os esporos. Uma alíquota de 100 µl da suspensão de esporo foi diluída em um volume
25 final de 5 ml com 0,01% de TWEEN® 80. Com o auxílio de um hemocitômetro, a concentração de esporo foi determinada e diluída em uma concentração final de 0,1 esporos por microlitro. Uma alíquota de 200 µl da diluição de esporo foi espalhada sobre 150 mm de placas de meio mínimo e incubada durante 2-3 dias a 34°C. As colônias emergindo foram excisadas

dentre as placas e transferidas para placas de PDA e incubadas durante 3 dias a 34°C. Então os esporos foram espalhados através das placas e incubados novamente durante 5 dias a 34°C.

5 As placas de esporos confluentes foram lavadas com 3 ml de
0,01% de TWEEN® 80 e a suspensão de esporo foi usada para inocular 25 ml
de meio MDU2BP em 125 ml de frascos de agitação de vidro. As culturas de
esporos únicas foram incubadas a 34°C com agitação constante a 200 rpm.
Após 5 dias, uma alíquota de 1 ml de cada cultura foi centrifugada a 12.000 x
g e seus sobrenadantes coletados. Cinco µl de cada sobrenadante foram
10 misturados com um volume igual de tampão de carga 2X (10% de beta-
mercaptoetanol) e carregados sobre um gel de SDS-PAGE 1,5 mm 8%-16%
Tris-glicina e coloridos com coloração de proteína BIO-SAFE® Commassie
Blue G250. Os perfis de SDS-PAGE dos caldos de cultura mostraram que
15 todos os oito transformantes foram capazes de expressar a proteína de fusão
de beta-glucosidase em muitos níveis elevados e uma das culturas designada
JaL355AIIo47 de *Aspergillus oryzae* produziu o melhor rendimento.

A invenção descrita e reivindicada aqui não é limitada em
escopo pelos aspectos específicos aqui descritos, porque estes aspectos se
destinam a ser ilustrações de vários aspectos da invenção. Quaisquer aspectos
20 equivalentes se destinam a estar dentro do escopo da invenção. De fato, várias
modificações da invenção além das mostradas e aqui descritas serão evidentes
para os versados na arte da descrição acima. Estas modificações também se
destinam a estar dentro do escopo das reivindicações anexas. No caso de
conflito, a presente descrição incluindo as definições será o controle.

25 Várias referências são citadas aqui, cujas descrições são
incorporadas por referência em suas totalidades

LISTAGEM DE SEQUÊNCIA

<110> Novozymes, Inc.
 <120> Métodos para aumentar secreção de polipeptídeos tendo atividade biológica
 <130> 10916.204-WO
 <150> US 60/832,511
 <151> 2006-07-21
 <160> 76
 <170> PatentIn version 3.4
 <210> 1
 <211> 923
 <212> DNA
 <213> Humicola insolens
 <400> 1

```

atgCGTtCct cccccctcct cCGtccGcc gTtGtggccg cCctgCCggt gTtggccott      60
gCCgctgatg gCaggtccac cGctactgg gactgtgca agccttctg cggctgggoc      120
aagaaggctc ccgtgaacca gcctgtcttt tCctgcaacg ccaacttcca gcgtatcacg      180
gacttcgacg ccaagtccgg ctgcgagccg ggcgggtgctg cctactctgt gcCCgaccag      240
accccatggg ctgtgaacga cGacttcGcg ctcggttttg ctgccacctc tattgCCggc      300
agcaatgagg cgggctggtg ctgcgcctgc tacgagctca ccttcacatc cggTcctgTt      360
gctggcaaga agatggtcgt ccagtcacc agcactggcg gtgatcttg cagcaaccac      420
ttcgatctca acatccccgg cggcggcgtc ggeatcttcg acggatgcac tccccagTtc      480
ggcggTctgc cgggccagcg ctacggcggc atctcgtccc gcaacgagtg cGatcggTtc      540
cccGacgccc tcaagcccgg ctgctactgg cGcttcgact ggttcaagaa cGccgacaat      600
ccgagcttca gcttccgtca ggtccagTgc ccagccgagc tcgtcGctcg caccggatgc      660
cGccgcaacg acgacggcaa cttccctgcc gtccagatcc cctccagcag caccagctct      720
cGgtcaacc agctaccag caccagcacc ccaccacctc gagcccgcca      780
gtccagccta cGactcccag cggctgcact gctgagaggt gggctcagtg cggcggcaat      840
ggctggagcg gctgcaccac ctgcgtcGct ggcagcactt gcaCGaagat taatgactgg      900
taccatcagt gcctgtagaa ttc

```

<210> 2

<211> 305

<212> PRT

<213> Humicola insolens

<400> 2

```

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro
1          5          10          15
Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
20        25        30
Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
35        40        45
Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
50        55        60
Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
65        70        75        80
Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
85        90        95
Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
100       105       110
Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
115       120       125
Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
130       135       140
Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
145       150       155       160
Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
165       170       175
Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
180       185       190
Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
195       200       205

```

Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
 210 215 220
 Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser
 225 230 235 240
 Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Ser Thr Ser Thr Thr
 245 250 255
 Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu
 260 265 270
 Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys
 275 280 285
 Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys
 290 295 300

Leu
 305

<210> 3

<211> 1188

<212> DNA

<213> Myceliophthora thermophila

<400> 3

```
cgacttgaaa cgccccaaat gaagtcctcc atcctcgcca gcgtcttcgc cacgggcgcc      60
gtggctcaaaa gtggctccgtg gcagcaatgt ggtggcatcg gatggcaagg atcgaccgac      120
tgtgtgtcgg gctaccactg cgtctaccag aacgattggt acagccagtg cgtgcctggc      180
gcggcgctcga caacgctgca gacatcgacc acgtccaggc ccaccgccac cagcaccgcc      240
cctccgctcgt ccaccacctc gcctagcaag ggcaagctga agtggctcgg cagcaacgag      300
tcgggcgccg agttcgggga gggcaattac cccggcctct ggggcaagca cttcatcttc      360
ccgtcgactt cggcgattca gacgctcatc aatgatggat acaacatctt cgggatcgac      420
ttctcgatgg agcgtctggt gcccaaccag ttgacgtcgt ccttcgacca gggttacctc      480
cgcaacctga ccgaggtggt caacttcgtg acgaacgcgg gcaagtacgc cgtcctggac      540
ccgcacaact acggccggta ctacggcaac atcatcacgg acacgaacgc gttccggacc      600
ttctggacca acctggccaa gcagttcgcc tccaactcgc tcgtcatctt cgacaccaac      660
aacgagtaca acacgatgga ccagaccctg gtgctcaacc tcaaccaggc cgccatcgac      720
ggcatccggg ccgcccggcg gacctcgag tacatcttcg tcgagggcaa cgctgggagc      780
ggggcctgga gctggaacac gaccaacacc aacatggccg ccctgacgga cccgcagaac      840
aagatcgtgt acgagatgca ccagtacctc gactcggaca gctcgggcac ccacgccgag      900
tgcgtcagca gcaccatcgg cgcccagcg gtcgctggag ccaaccagtg gctcccgcgc      960
aacggcaagc tcggcgtcct cggcgagttc ccgggcggcg ccaacgccgt ctgccagcag     1020
gccgtcaccg gcctcctcga ccacctccag gacaacagcg acgtctggct gggtgccctc     1080
tggtgggccc ccggtccctg gtggggcgac tacatgtact cgttcgagcc tccttcggggc     1140
accggctatg tcaactaaa ctcgatcttg aagaagtact tgccgtaa                       1188
```

<210> 4

<211> 389

<212> PRT

<213> Myceliophthora thermophila

<400> 4

```
Met Lys Ser Ser Ile Leu Ala Ser Val Phe Ala Thr Gly Ala Val Ala
 1 5 10 15
Gln Ser Gly Pro Trp Gln Gln Cys Gly Gly Ile Gly Trp Gln Gly Ser
 20 25 30
Thr Asp Cys Val Ser Gly Tyr His Cys Val Tyr Gln Asn Asp Trp Tyr
 35 40 45
Ser Gln Cys Val Pro Gly Ala Ala Ser Thr Thr Leu Gln Thr Ser Thr
 50 55 60
Thr Ser Arg Pro Thr Ala Thr Ser Thr Ala Pro Ser Ser Thr Thr
 65 70 75 80
Ser Pro Ser Lys Gly Lys Leu Lys Trp Leu Gly Ser Asn Glu Ser Gly
 85 90 95
Ala Glu Phe Gly Glu Gly Asn Tyr Pro Gly Leu Trp Gly Lys His Phe
 100 105 110
Ile Phe Pro Ser Thr Ser Ala Ile Gln Thr Leu Ile Asn Asp Gly Tyr
 115 120 125
Asn Ile Phe Arg Ile Asp Phe Ser Met Glu Arg Leu Val Pro Asn Gln
 130 135 140
Leu Thr Ser Ser Phe Asp Gln Gly Tyr Leu Arg Asn Leu Thr Glu Val
```


1				5					10					15	
Pro	Val	Phe	Ser	Val	Ala	Val	Trp	Gly	Gln	Cys	Gly	Gly	Ile	Gly	Phe
			20					25					30		
Ser	Gly	Ser	Thr	Val	Cys	Asp	Ala	Gly	Ala	Gly	Cys	Val	Lys	Leu	Asn
		35				40					45				
Asp	Tyr	Tyr	Ser	Gln	Cys	Gln	Pro	Gly	Ala	Pro	Thr	Ala	Thr	Ser	Ala
	50					55					60				
Ala	Pro	Ser	Ser	Asn	Ala	Pro	Ser	Gly	Thr	Ser	Thr	Ala	Ser	Ala	Pro
65				70						75					80
Ser	Ser	Ser	Leu	Cys	Ser	Gly	Ser	Arg	Thr	Pro	Phe	Gln	Phe	Phe	Gly
				85				90						95	
Val	Asn	Glu	Ser	Gly	Ala	Glu	Phe	Gly	Asn	Leu	Asn	Ile	Pro	Gly	Val
			100					105					110		
Leu	Gly	Thr	Asp	Tyr	Thr	Trp	Pro	Ser	Pro	Ser	Ser	Ile	Asp	Phe	Phe
		115					120					125			
Met	Gly	Lys	Gly	Met	Asn	Thr	Phe	Arg	Ile	Pro	Phe	Leu	Met	Glu	Arg
	130					135					140				
Leu	Val	Pro	Pro	Ala	Thr	Gly	Ile	Thr	Gly	Pro	Leu	Asp	Gln	Thr	Tyr
145				150						155					160
Leu	Gly	Gly	Leu	Gln	Thr	Ile	Val	Asn	Tyr	Ile	Thr	Gly	Lys	Gly	Gly
				165				170						175	
Phe	Ala	Leu	Ile	Asp	Pro	His	Asn	Phe	Met	Ile	Tyr	Asn	Gly	Gln	Thr
			180					185					190		
Ile	Ser	Ser	Thr	Ser	Asp	Phe	Gln	Lys	Phe	Trp	Gln	Asn	Leu	Ala	Gly
		195				200					205				
Val	Phe	Lys	Ser	Asn	Ser	His	Val	Ile	Phe	Asp	Val	Met	Asn	Glu	Pro
	210					215					220				
His	Asp	Ile	Pro	Ala	Gln	Thr	Val	Phe	Gln	Leu	Asn	Gln	Ala	Ala	Val
225					230					235					240
Asn	Gly	Ile	Arg	Ala	Ser	Gly	Ala	Thr	Ser	Gln	Leu	Ile	Leu	Val	Glu
				245					250					255	
Gly	Thr	Ser	Trp	Thr	Gly	Ala	Trp	Thr	Trp	Thr	Thr	Ser	Gly	Asn	Ser
		260						265					270		
Asp	Ala	Phe	Gly	Ala	Ile	Lys	Asp	Pro	Asn	Asn	Asn	Val	Ala	Ile	Gln
		275					280					285			
Met	His	Gln	Tyr	Leu	Asp	Ser	Asp	Gly	Ser	Gly	Thr	Ser	Gln	Thr	Cys
	290					295					300				
Val	Ser	Pro	Thr	Ile	Gly	Ala	Glu	Arg	Leu	Gln	Ala	Ala	Thr	Gln	Trp
305					310					315					320
Leu	Lys	Gln	Asn	Asn	Leu	Lys	Gly	Phe	Leu	Gly	Glu	Ile	Gly	Ala	Gly
				325					330					335	
Ser	Asn	Ser	Ala	Cys	Ile	Ser	Ala	Val	Gln	Gly	Ala	Leu	Cys	Ser	Met
			340					345					350		
Gln	Gln	Ser	Gly	Val	Trp	Leu	Gly	Ala	Leu	Trp	Trp	Ala	Ala	Gly	Pro
		355					360					365			
Trp	Trp	Gly	Asp	Tyr	Tyr	Gln	Ser	Ile	Glu	Pro	Pro	Ser	Gly	Pro	Ala
	370					375					380				
Val	Ser	Ala	Ile	Leu	Pro	Gln	Ala	Leu	Leu	Pro	Phe	Ala			
385					390					395					

<210> 7

<211> 1303

<212> DNA

<213> BASIDIOMYCETE CBS 495.95

<400> 7

ggaaagcgtc	agtatggtga	aatttgcgct	tgtggcaact	gtcggcgcaa	tcttgagcgc	60
ttctgcggcc	aatgcggtt	ctatctacca	gcaatgtgga	ggcattggat	ggtctgggtc	120
cactgtttgc	gacgccggtc	tcgcttgcgt	tatcctcaat	gcgtactact	ttcagtgcct	180
gacgcccgcc	gcgggccaga	caacgacggg	ctcgggcgca	ccggcgtcaa	catcaacctc	240
tcaactcaacg	gtcactacgg	ggagctcaca	ctcaacaacc	gggacgacgg	cgacgaaaac	300
aactaccact	ccgtcgacca	ccacgaccct	accgcaccatc	tctgtgtctg	gtcgcgtctg	360
ctctggctcc	aggacgaagt	tcaagttctt	cggtgtgaat	gaaagcggcg	ccgaattcgg	420
gaacactgct	tggccagggc	agctcgggaa	agactataca	tggccttcgc	ctagcagcgt	480
ggactacttc	atgggggctg	gattcaatac	attccgtatc	accttcttga	tggagcgtat	540

gagccctccg gctaccggac tcaactggccc attcaaccag acgtacctgt cgggectcac 600
caccattgtc gactacatca cgaacaaagg aggatacgtt cttattgacc cccacaactt 660
catgcgttac aacaacggca taatcagcag cacatctgac ttcgcgactt ggtggagcaa 720
tttggccact gtattcaaat ccacgaagaa cgccatcttc gacatccaga acgagccgta 780
cggaatcgat ggcgagaccg tatacgaact gaatcaagct gccatcaatt cgatccgcg 840
cgctggcgct acgtcacagt tgattctggt tgaaggaacg tcatacactg gagcttggac 900
gtgggtctcg tccggaaacg gagctgcttt cgcggccggt acggatcctt acaacaacac 960
ggcaattgaa atgcaccaat acctcgacag cgacggttct gggacaaacg aagactgtgt 1020
ctctccacc attgggtcgc aacgtctcca agctgccact gcgtggctgc acaaacagg 1080
actcaaggga ttctcggag agacgggtgc tgggtcgaat tcccagtgca tcgacgccgt 1140
gttcgatgaa ctttctata tgcaacagca aggcggctcc tggatcgggt cactctgggt 1200
ggctgcggtt ccttgggtgg gcacgtacat ttactcgatt gaacctccga gcggtgccgc 1260
tatcccagaa gtccttctc agggctctgc tccattctc tag 1303
<210> 8
<211> 429
<212> PRT
<213> BASIDIOMYCETE CBS 495.95
<400> 8
Met Val Lys Phe Ala Leu Val Ala Thr Val Gly Ala Ile Leu Ser Ala
1 5 10 15
Ser Ala Ala Asn Ala Ala Ser Ile Tyr Gln Gln Cys Gly Gly Ile Gly
20 25 30
Trp Ser Gly Ser Thr Val Cys Asp Ala Gly Leu Ala Cys Val Ile Leu
35 40 45
Asn Ala Tyr Tyr Phe Gln Cys Leu Thr Pro Ala Ala Gly Gln Thr Thr
50 55 60
Thr Gly Ser Gly Ala Pro Ala Ser Thr Ser Thr Ser His Ser Thr Val
65 70 75 80
Thr Thr Gly Ser Ser His Ser Thr Thr Thr Gly Thr Thr Ala Thr Lys Thr
85 90 95
Thr Thr Thr Pro Ser Thr Thr Thr Thr Leu Pro Ala Ile Ser Val Ser
100 105 110
Gly Arg Val Cys Ser Gly Ser Arg Thr Lys Phe Lys Phe Phe Gly Val
115 120 125
Asn Glu Ser Gly Ala Glu Phe Gly Asn Thr Ala Trp Pro Gly Gln Leu
130 135 140
Gly Lys Asp Tyr Thr Trp Pro Ser Pro Ser Ser Val Asp Tyr Phe Met
145 150 155 160
Gly Ala Gly Phe Asn Thr Phe Arg Ile Thr Phe Leu Met Glu Arg Met
165 170 175
Ser Pro Pro Ala Thr Gly Leu Thr Gly Pro Phe Asn Gln Thr Tyr Leu
180 185 190
Ser Gly Leu Thr Thr Ile Val Asp Tyr Ile Thr Asn Lys Gly Gly Tyr
195 200 205
Ala Leu Ile Asp Pro His Asn Phe Met Arg Tyr Asn Asn Gly Ile Ile
210 215 220
Ser Ser Thr Ser Asp Phe Ala Thr Trp Trp Ser Asn Leu Ala Thr Val
225 230 235 240
Phe Lys Ser Thr Lys Asn Ala Ile Phe Asp Ile Gln Asn Glu Pro Tyr
245 250 255
Gly Ile Asp Ala Gln Thr Val Tyr Glu Leu Asn Gln Ala Ala Ile Asn
260 265 270
Ser Ile Arg Ala Ala Gly Ala Thr Ser Gln Leu Ile Leu Val Glu Gly
275 280 285
Thr Ser Tyr Thr Gly Ala Trp Thr Trp Val Ser Ser Gly Asn Gly Ala
290 295 300
Ala Phe Ala Ala Val Thr Asp Pro Tyr Asn Asn Thr Ala Ile Glu Met
305 310 315 320
His Gln Tyr Leu Asp Ser Asp Gly Ser Gly Thr Asn Glu Asp Cys Val
325 330 335
Ser Ser Thr Ile Gly Ser Gln Arg Leu Gln Ala Ala Thr Ala Trp Leu
340 345 350
Gln Gln Thr Gly Leu Lys Gly Phe Leu Gly Glu Thr Gly Ala Gly Ser

130						135						140					
Glu Tyr Ile Asp Lys Ile Val Ser Ile Leu Lys Ala His Pro Asn Thr																	
145						150						155					160
Ala Phe Ala Leu Val Ile Glu Pro Asp Ser Leu Pro Asn Leu Val Thr																	
				165								170					175
Asn Ser Asn Leu Asp Thr Cys Ser Ser Ser Ala Ser Gly Tyr Arg Glu																	
			180														190
Gly Val Ala Tyr Ala Leu Lys Asn Leu Asn Leu Pro Asn Val Ile Met																	
			195					200									205
Tyr Leu Asp Ala Gly His Gly Gly Trp Leu Gly Trp Asp Ala Asn Leu																	
			210					215									220
Gln Pro Gly Ala Gln Glu Leu Ala Lys Ala Tyr Lys Asn Ala Gly Ser																	
			225					230									240
Pro Lys Gln Leu Arg Gly Phe Ser Thr Asn Val Ala Gly Trp Asn Ser																	
			245														255
Trp Asp Gln Ser Pro Gly Glu Phe Ser Gln Ala Ser Asp Ala Lys Tyr																	
			260														270
Asn Lys Cys Gln Asn Glu Lys Ile Tyr Val Ser Thr Phe Gly Ser Ala																	
			275														285
Leu Gln Ser Ala Gly Met Pro Asn His Ala Ile Val Asp Thr Gly Arg																	
			290					295									300
Asn Gly Val Thr Gly Leu Arg Lys Glu Trp Gly Asp Trp Cys Asn Val																	
			305					310									320
Asn Gly Ala Gly Phe Gly Val Arg Pro Thr Ser Asn Thr Gly Leu Glu																	
			325														335
Leu Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly																	
			340														350
Thr Ser Asp Ser Ser Ser Pro Arg Tyr Asp Ser Phe Cys Gly Lys Asp																	
			355														365
Asp Ala Phe Lys Pro Ser Pro Glu Ala Gly Thr Trp Asn Glu Ala Tyr																	
			370					375									380
Phe Glu Met Leu Leu Lys Asn Ala Val Pro Ser Phe																	
			385														395

<210> 11

<211> 1203

<212> DNA

<213> *Thielavia terrestris*

<400> 11

atgaagtacc	tcaacctcct	cgcagctctc	ctcgcgctcg	ctcctctctc	cctcgcctgca	60
cccagcatcg	aggccagaca	gtcgaacgtc	aaccataca	tccgcaagag	cccgcctcgtt	120
attaggtcgt	acgccccaaa	gcttgaggag	accgtcagga	ccttccagca	acgtggcgac	180
cagctcaacg	ctgcgaggac	acggacgggtg	cagaacgttg	cgactttcgc	ctggatctcg	240
gataccaatg	gtattggagc	cattcgacct	ctcatccaag	atgctctcgc	ccagcaggct	300
cgcactggac	agaaggtcat	cgtccaaatc	gtcgtctaca	acctcccaga	tcgcgactgc	360
tctgccaacg	cctcgactgg	agagttcacc	gtaggaaacg	acggtctcaa	ccgatacaag	420
aactttgtca	acaccatcgc	ccgcgagctc	tcgactgctg	acgctgacaa	gtcccacttt	480
gcctctctcc	tcgaaccoga	cgcacttgcc	aacctcgtca	ccaacgcgaa	tgcccccagg	540
tgccgaatcg	ccgctccccg	ttacaaggag	ggatcgcct	acaccctcgc	caccttgctc	600
aagcccaacg	tcgacgtcta	catcgacgcc	gccaacgggtg	gctggctcgg	ctggaacgac	660
aacctccgcc	ccttcgccga	actcttcaag	gaagtctacg	acctcgcctg	ccgcatcaac	720
cccaacgcca	aggtcgcggg	cgcccccgtc	aacgtctcca	actacaacca	gtaccgcgct	780
gaagtccgcg	agcccttcac	cgagtggaag	gacgcctggg	acgagagccg	ctacgtcaac	840
gtctcacc	cgcacctcaa	cgccgtcggc	ttctccgcgc	acttcatcgt	tgaccaggga	900
cgcggtggca	agggcggtat	caggacggag	tggggccagt	gggtgcaactg	taggaacgct	960
gggttcggta	tcaggcctac	tgcggatcag	ggcgtgctcc	agaaccgaa	tgtggatgcg	1020
attgtgtggg	ttaagccggg	tggagagtcg	gatggcaacg	gtgatttgaa	ctcgaacagg	1080
tatgatccta	cgtgcaggag	tccggtggcg	catgttcccg	ctcctgaggg	tggccagtgg	1140
ttcaacgagt	atggtgttaa	cctcgttttg	aacgctaacc	cccctcttga	gcctacctgg	1200
taa						1203

<210> 12

<211> 400

<212> PRT

<213> *Thielavia terrestris*

<400> 12

Met Lys Tyr Leu Asn Leu Leu Ala Ala Leu Leu Ala Val Ala Pro Leu
 1 5 10 15
 Ser Leu Ala Ala Pro Ser Ile Glu Ala Arg Gln Ser Asn Val Asn Pro
 20 25 30
 Tyr Ile Gly Lys Ser Pro Leu Val Ile Arg Ser Tyr Ala Gln Lys Leu
 35 40 45
 Glu Glu Thr Val Arg Thr Phe Gln Gln Arg Gly Asp Gln Leu Asn Ala
 50 55 60
 Ala Arg Thr Arg Thr Val Gln Asn Val Ala Thr Phe Ala Trp Ile Ser
 65 70 75 80
 Asp Thr Asn Gly Ile Gly Ala Ile Arg Pro Leu Ile Gln Asp Ala Leu
 85 90 95
 Ala Gln Gln Ala Arg Thr Gly Gln Lys Val Ile Val Gln Ile Val Val
 100 105 110
 Tyr Asn Leu Pro Asp Arg Asp Cys Ser Ala Asn Ala Ser Thr Gly Glu
 115 120 125
 Phe Thr Val Gly Asn Asp Gly Leu Asn Arg Tyr Lys Asn Phe Val Asn
 130 135 140
 Thr Ile Ala Arg Glu Leu Ser Thr Ala Asp Ala Asp Lys Leu His Phe
 145 150 155 160
 Ala Leu Leu Leu Glu Pro Asp Ala Leu Ala Asn Leu Val Thr Asn Ala
 165 170 175
 Asn Ala Pro Arg Cys Arg Ile Ala Ala Pro Ala Tyr Lys Glu Gly Ile
 180 185 190
 Ala Tyr Thr Leu Ala Thr Leu Ser Lys Pro Asn Val Asp Val Tyr Ile
 195 200 205
 Asp Ala Ala Asn Gly Gly Trp Leu Gly Trp Asn Asp Asn Leu Arg Pro
 210 215 220
 Phe Ala Glu Leu Phe Lys Glu Val Tyr Asp Leu Ala Arg Arg Ile Asn
 225 230 235 240
 Pro Asn Ala Lys Val Arg Gly Val Pro Val Asn Val Ser Asn Tyr Asn
 245 250 255
 Gln Tyr Arg Ala Glu Val Arg Glu Pro Phe Thr Glu Trp Lys Asp Ala
 260 265 270
 Trp Asp Glu Ser Arg Tyr Val Asn Val Leu Thr Pro His Leu Asn Ala
 275 280 285
 Val Gly Phe Ser Ala His Phe Ile Val Asp Gln Gly Arg Gly Gly Lys
 290 295 300
 Gly Gly Ile Arg Thr Glu Trp Gly Gln Trp Cys Asn Val Arg Asn Ala
 305 310 315 320
 Gly Phe Gly Ile Arg Pro Thr Ala Asp Gln Gly Val Leu Gln Asn Pro
 325 330 335
 Asn Val Asp Ala Ile Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly
 340 345 350
 Thr Ser Asp Leu Asn Ser Asn Arg Tyr Asp Pro Thr Cys Arg Ser Pro
 355 360 365
 Val Ala His Val Pro Ala Pro Glu Ala Gly Gln Trp Phe Asn Glu Tyr
 370 375 380
 Val Val Asn Leu Val Leu Asn Ala Asn Pro Pro Leu Glu Pro Thr Trp
 385 390 395 400

<210> 13

<211> 1501

<212> DNA

<213> *Thielavia terrestris*

<400> 13

gccgttggtca agatgggcca gaagacgctg cacggattcg cgcgccaggc tttggccggtt 60
 ctcccctttg tgaaggctca gcagcccggc aacttcacgc cggagggtgca cccgcaactg 120
 ccaacgtgga agtgcacgac cgccggcggc tgcgttcacg aggacacttc ggtggtgctc 180
 gactggaact accgttgat ccacaatgcc gacggcaccg cctcgtgcac gacgtccagc 240
 ggggtcgacc acacgctgtg tccagatgag gcgacctgcg cgaagaactg cttcgtggaa 300
 ggcgtcaact acacgagcag cggtgtcacc acatccggca gttcgtgac gatgaggcag 360
 tatttcaagg ggagcaacgg gcagaccaac agcgtttcgc ctcgtctcta cctgctcggc 420

```

tcggatggaa actacgtaat gctcaagctg ctccggccagg agctgagctt cgatgtcgat 480
ctctccacgc tcccctgctg cgagaacggc gcgctgtacc tgtecgagat ggacgcgacc 540
ggtggcagga accagtataa caccggcggg gccaaactacg gctcgggcta ctgtgacgcc 600
cagtgtcccg tgcagacgtg gatgaacggc acgctgaaca ccaacgggca gggctactgc 660
tgcaacgaga tggacatcct cgaggccaac tcccgcgcca acgcatgac acctcacccc 720
tgcgccaacg gcagctgcca caagagcggg tgcggactca acccctacgc cgagggtac 780
aagagctact acggaccggg cctcacgggt gacacgtcga agcccttcac catcattacc 840
cgcttcatca ccgacgacgg caccgaccagc ggcaccctca accagatcca gcggatctat 900
gtgcagaatg gcaagacggg cgcgtcggct cgcgtccggag gcgacatcat cacggcatcc 960
ggctgcacct cggcccagge gttcggcggg ctggccaaca tgggcgcggc gcttgacgg 1020
ggcatggtgc tgaccttcag catctggaac gacgctgggg gctacatgaa ctggctcgac 1080
agcggcaaca acggcccgtg cagcagcacc gagggcaacc cgtccaacat cctggccaac 1140
taccgggaca cccacgtggt cttctccaac atccgctggg gagacatcgg ctcgacggtc 1200
caggtctcgg gaggcggcaa cggcggtcgc accaccacca cgtcgaccac cacgctgagg 1260
acctcgacca cgaccaccac caccgccccg acggccactg ccacgcaactg gggacaatgc 1320
ggcggaatcg gggtagtca accgcctcct gcattctggt gaggaagtta actaacgtgg 1380
cctacgcagt ggactggacc gaccgtctgc gaatcgccgt acgcatgcaa ggagctgaac 1440
ccctggtact accagtgcct ctaaagtatt gcagtgaaagc catactccgt gctcggcatg 1500

```

g

<210> 14

<211> 464

<212> PRT

<213> *Thielavia terrestris*

<400> 14

```

Met Gly Gln Lys Thr Leu His Gly Phe Ala Ala Thr Ala Leu Ala Val
1 5 10 15
Leu Pro Phe Val Lys Ala Gln Gln Pro Gly Asn Phe Thr Pro Glu Val
20 25 30
His Pro Gln Leu Pro Thr Trp Lys Cys Thr Thr Ala Gly Gly Cys Val
35 40 45
Gln Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His
50 55 60
Asn Ala Asp Gly Thr Ala Ser Cys Thr Thr Ser Ser Gly Val Asp His
65 70 75 80
Thr Leu Cys Pro Asp Glu Ala Thr Cys Ala Lys Asn Cys Phe Val Glu
85 90 95
Gly Val Asn Tyr Thr Ser Ser Gly Val Thr Thr Ser Gly Ser Ser Leu
100 105 110
Thr Met Arg Gln Tyr Phe Lys Gly Ser Asn Gly Gln Thr Asn Ser Val
115 120 125
Ser Pro Arg Leu Tyr Leu Leu Gly Ser Asp Gly Asn Tyr Val Met Leu
130 135 140
Lys Leu Leu Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Thr Leu
145 150 155 160
Pro Cys Gly Glu Asn Gly Ala Leu Tyr Leu Ser Glu Met Asp Ala Thr
165 170 175
Gly Gly Arg Asn Gln Tyr Asn Thr Gly Gly Ala Asn Tyr Gly Ser Gly
180 185 190
Tyr Cys Asp Ala Gln Cys Pro Val Gln Thr Trp Met Asn Gly Thr Leu
195 200 205
Asn Thr Asn Gly Gln Gly Tyr Cys Cys Asn Glu Met Asp Ile Leu Glu
210 215 220
Ala Asn Ser Arg Ala Asn Ala Met Thr Pro His Pro Cys Ala Asn Gly
225 230 235 240
Ser Cys Asp Lys Ser Gly Cys Gly Leu Asn Pro Tyr Ala Glu Gly Tyr
245 250 255
Lys Ser Tyr Tyr Gly Pro Gly Leu Thr Val Asp Thr Ser Lys Pro Phe
260 265 270
Thr Ile Ile Thr Arg Phe Ile Thr Asp Asp Gly Thr Thr Ser Gly Thr
275 280 285
Leu Asn Gln Ile Gln Arg Ile Tyr Val Gln Asn Gly Lys Thr Val Ala
290 295 300
Ser Ala Ala Ser Gly Gly Asp Ile Ile Thr Ala Ser Gly Cys Thr Ser

```


				85						90				95			
His	Gly	Val	Ala	Thr	Lys	Gly	Asp	Ala	Leu	Thr	Leu	His	Gln	Trp	Leu		
			100					105					110				
Lys	Gly	Ala	Asp	Gly	Thr	Tyr	Arg	Thr	Val	Ser	Pro	Arg	Val	Tyr	Leu		
		115					120					125					
Leu	Gly	Glu	Asp	Gly	Lys	Asn	Tyr	Glu	Asp	Phe	Lys	Leu	Leu	Asn	Ala		
		130				135					140						
Glu	Leu	Ser	Phe	Asp	Val	Asp	Val	Ser	Gln	Leu	Val	Cys	Gly	Met	Asn		
145					150					155					160		
Gly	Ala	Leu	Tyr	Phe	Ser	Glu	Met	Glu	Met	Asp	Gly	Gly	Arg	Ser	Pro		
				165				170						175			
Leu	Asn	Pro	Ala	Gly	Ala	Thr	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp	Ala	Gln		
			180					185					190				
Cys	Pro	Lys	Leu	Asp	Phe	Ile	Asn	Gly	Glu	Leu	Asn	Thr	Asn	His	Thr		
		195					200					205					
Tyr	Gly	Ala	Cys	Cys	Asn	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	Ala	Leu		
	210					215					220						
Ala	Gln	Ala	Leu	Thr	Pro	His	Pro	Cys	Asn	Ala	Thr	Arg	Val	Tyr	Lys		
225					230					235					240		
Cys	Asp	Thr	Ala	Asp	Glu	Cys	Gly	Gln	Pro	Val	Gly	Val	Cys	Asp	Glu		
				245				250						255			
Trp	Gly	Cys	Ser	Tyr	Asn	Pro	Ser	Asn	Phe	Gly	Val	Lys	Asp	Tyr	Tyr		
			260					265						270			
Gly	Arg	Asn	Leu	Thr	Val	Asp	Thr	Asn	Arg	Lys	Phe	Thr	Val	Thr	Thr		
		275					280					285					
Gln	Phe	Val	Thr	Ser	Asn	Gly	Arg	Ala	Asp	Gly	Glu	Leu	Thr	Glu	Ile		
	290					295					300						
Arg	Arg	Leu	Tyr	Val	Gln	Asp	Gly	Val	Val	Ile	Gln	Asn	His	Ala	Val		
305					310					315					320		
Thr	Ala	Gly	Gly	Ala	Thr	Tyr	Asp	Ser	Ile	Thr	Asp	Gly	Phe	Cys	Asn		
				325				330						335			
Ala	Thr	Ala	Thr	Trp	Thr	Gln	Gln	Arg	Gly	Gly	Leu	Ala	Arg	Met	Gly		
			340					345						350			
Glu	Ala	Ile	Gly	Arg	Gly	Met	Val	Leu	Ile	Phe	Ser	Leu	Trp	Val	Asp		
		355					360					365					
Asn	Gly	Gly	Phe	Met	Asn	Trp	Leu	Asp	Ser	Gly	Asn	Ala	Gly	Pro	Cys		
	370					375					380						
Asn	Ala	Thr	Glu	Gly	Asp	Pro	Ala	Leu	Ile	Leu	Gln	Gln	His	Pro	Asp		
385					390					395					400		
Ala	Ser	Val	Thr	Phe	Ser	Asn	Ile	Arg	Trp	Gly	Glu	Ile	Gly	Ser	Thr		
				405				410						415			
Tyr	Lys	Ser	Glu	Cys	Ser	His											
			420														

<210> 17

<211> 1011

<212> DNA

<213> *Thielavia terrestris*

<400> 17

atgaccctac	ggctccctgt	catcagcctg	ctggcctcgc	tggcagcagg	cgccgtcgtc		60
gtcccacggg	cggagtttca	ccccctctc	ccgacttgga	aatgcacgac	ctccggggggc		120
tgcgtgcagc	agaacaccag	cgctcgtcctg	gaccgtgact	cgaagtacgc	cgcacacagc		180
gccggctcgc	ggacggaatc	ggattacgcg	gcaatgggag	tgtccacttc	gggcaatgcc		240
gtgacgctgt	accactacgt	caagaccaac	ggcaccctgc	tccccgcttc	gccgcgcctc		300
tacctcctgg	gcgcggaagg	caagtacgtg	cttatggacc	tcctcaacca	ggagctgtcg		360
gtggacgtcg	acttctcggc	gctgcccgtgc	ggcgagaacg	gggccttcta	cctgtccgag		420
atggcgcgcg	acgggcgggg	cgacgcgggg	gcgggagcag	ggtaactgca	cgcgagtgcc		480
cagggtact	gctgcaacga	gatggacatc	ctcgaggcca	actcgatggc	gacggccatg		540
acgccgcacc	cgtgcaaggg	caacaactgc	gaccgcagcg	gctgoggcta	caacccttac		600
gccagcggcc	agcgcggctt	ctacggggccc	ggcaagacgg	tcgacacgag	caagcccttc		660
accgtcgtca	cgcagttcgc	cgccagcggc	ggcaagctga	cccagatcac	ccgcaagtac		720
atccagaacg	gccgggagat	cggcggcggc	ggcaccatct	ccagctgcgg	ctccgagtct		780
tcgacgggcg	gctgaccggg	catgggcgag	gcgctggggc	gcggaatggt	gctggccatg		840
agcatctgga	acgacgcggc	ccaggagatg	gcatggctcg	atgccggcaa	caacggccct		900

tgcgccagtg gccagggcag cccgtccgtc attcagtcgc agcatcccga caccacgtc 960
 gtctttctcca acatcaggtg gggcgacatc gggcttacca cgaagaacta g 1011

<210> 18

<211> 336

<212> PRT

<213> Thielavia terrestris

<400> 18

Met Thr Leu Arg Leu Pro Val Ile Ser Leu Leu Ala Ser Leu Ala Ala
 1 5 10 15
 Gly Ala Val Val Val Pro Arg Ala Glu Phe His Pro Pro Leu Pro Thr
 20 25 30
 Trp Lys Cys Thr Thr Ser Gly Gly Cys Val Gln Gln Asn Thr Ser Val
 35 40 45
 Val Leu Asp Arg Asp Ser Lys Tyr Ala Ala His Ser Ala Gly Ser Arg
 50 55 60
 Thr Glu Ser Asp Tyr Ala Ala Met Gly Val Ser Thr Ser Gly Asn Ala
 65 70 75 80
 Val Thr Leu Tyr His Tyr Val Lys Thr Asn Gly Thr Leu Val Pro Ala
 85 90 95
 Ser Pro Arg Ile Tyr Leu Leu Gly Ala Asp Gly Lys Tyr Val Leu Met
 100 105 110
 Asp Leu Leu Asn Gln Glu Leu Ser Val Asp Val Asp Phe Ser Ala Leu
 115 120 125
 Pro Cys Gly Glu Asn Gly Ala Phe Tyr Leu Ser Glu Met Ala Ala Asp
 130 135 140
 Gly Arg Gly Asp Ala Gly Ala Gly Asp Gly Tyr Cys Asp Ala Gln Cys
 145 150 155 160
 Gln Gly Tyr Cys Cys Asn Glu Met Asp Ile Leu Glu Ala Asn Ser Met
 165 170 175
 Ala Thr Ala Met Thr Pro His Pro Cys Lys Gly Asn Asn Cys Asp Arg
 180 185 190
 Ser Gly Cys Gly Tyr Asn Pro Tyr Ala Ser Gly Gln Arg Gly Phe Tyr
 195 200 205
 Gly Pro Gly Lys Thr Val Asp Thr Ser Lys Pro Phe Thr Val Val Thr
 210 215 220
 Gln Phe Ala Ala Ser Gly Gly Lys Leu Thr Gln Ile Thr Arg Lys Tyr
 225 230 235 240
 Ile Gln Asn Gly Arg Glu Ile Gly Gly Gly Gly Thr Ile Ser Ser Cys
 245 250 255
 Gly Ser Glu Ser Ser Thr Gly Gly Leu Thr Gly Met Gly Glu Ala Leu
 260 265 270
 Gly Arg Gly Met Val Leu Ala Met Ser Ile Trp Asn Asp Ala Ala Gln
 275 280 285
 Glu Met Ala Trp Leu Asp Ala Gly Asn Asn Gly Pro Cys Ala Ser Gly
 290 295 300
 Gln Gly Ser Pro Ser Val Ile Gln Ser Gln His Pro Asp Thr His Val
 305 310 315 320
 Val Phe Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser Thr Thr Lys Asn
 325 330 335

<210> 19

<211> 1480

<212> DNA

<213> Cladorrhinum foecundissimum

<400> 19

gatccgaatt cctcctctcg ttcttttagtc acagaccaga catctgcca cgatggttca 60
 caagttcgcc ctctcaccg gcctcgccgc ctccctcgca tctgcccagc agatcggcac 120
 cgtcgtcccc gagtctcacc ccaagcttc caccaagcgc tgcactctcg cgggtggctg 180
 ccagaccgtc gacacctcca tegtcatcga cgcttccag cgtcccctcc acaagatcgg 240
 cgacccttcc actccttgcg tegtggcgg cctctctgtc cccgacgcca agtctctgcg 300
 tgagaactgc gcgctcgagg gtgtcgacta tgcctcctgg ggcatacaaga ccgagggcga 360
 cgccctaact ctcaaccagt ggatgccga cccggcgaac cctggccagt acaagacgac 420
 tactccccgt acttaccttg ttgctgagga cggcaagaac tacgaggatg tgaagctcct 480
 ggctaaggag atctcgtttg atgccgatgt cagcaacctt ccctgoggca tgaacgggtg 540

```

tttctacttg tctgagatgt tgatggatgg tggacgtggc gacctcaacc ctgctgggtgc 600
cgagtatggg accggttact gtgatgcgca gtgcttcaag ttggatttca tcaacggcga 660
ggccaacatc gaccaaagc acggcgctg ctgcaacgaa atggacattt tcaaatccaa 720
ctcgcgcgcc aagaccttcg tccccacc ctgcaacatc acgcaggtct acaagtgcga 780
aggcgaagac gagtgccggc agcccgtcgg cgtgtgcgac aagtgggggt gcggttcaa 840
cgagtacaaa tggggcgctg agtccttcta cggccggggc tcgcagttcg ccatcgactc 900
ctccaagaag ttcaccgtca ccacgcagtt cctgaccgac aacggcaagg aggacggcgt 960
cctcgtcgag atccgcccgt tgtggcacca ggtggcaag ctgatcaaga acaccgctat 1020
ccaggttgag gagaactaca gcacggactc ggtgagcacc gagttctgcg agaagactgc 1080
ttctttcacc atgcagcgcg gtggtctcaa ggcgatgggc gaggctatcg gtcgtggtat 1140
ggtgctgggt ttcagcatct gggcggatga ttcgggtttt atgaactggt tggatgcgga 1200
gggtaatggc ccttgcagcg cgactgaggg cgatccgaag gagattgtca agaataagcc 1260
ggatgctagg gttacgttct caaacattag gattggtgag gttggtagca cgtatgctcc 1320
gggtgggaag tgcggtgtta agagcagggg tgctaggggg cttactgctt cttagggggg 1380
gtgtgaagag aggaggaggt gttggtgggg gttggagatg ataattgggc gagatggtgt 1440
agagcgggtt ggttgatgat gaatacgttg aattggatgt 1480
<210> 20
<211> 440
<212> PRT
<213> Cladorrhinum foecundissimum
<400> 20
Met Val His Lys Phe Ala Leu Leu Thr Gly Leu Ala Ala Ser Leu Ala
1 5 10 15
Ser Ala Gln Gln Ile Gly Thr Val Val Pro Glu Ser His Pro Lys Leu
20 25 30
Pro Thr Lys Arg Cys Thr Leu Ala Gly Gly Cys Gln Thr Val Asp Thr
35 40 45
Ser Ile Val Ile Asp Ala Phe Gln Arg Pro Leu His Lys Ile Gly Asp
50 55 60
Pro Ser Thr Pro Cys Val Val Gly Gly Pro Leu Cys Pro Asp Ala Lys
65 70 75 80
Ser Cys Ala Glu Asn Cys Ala Leu Glu Gly Val Asp Tyr Ala Ser Trp
85 90 95
Gly Ile Lys Thr Glu Gly Asp Ala Leu Thr Leu Asn Gln Trp Met Pro
100 105 110
Asp Pro Ala Asn Pro Gly Gln Tyr Lys Thr Thr Thr Pro Arg Thr Tyr
115 120 125
Leu Val Ala Glu Asp Gly Lys Asn Tyr Glu Asp Val Lys Leu Leu Ala
130 135 140
Lys Glu Ile Ser Phe Asp Ala Asp Val Ser Asn Leu Pro Cys Gly Met
145 150 155 160
Asn Gly Ala Phe Tyr Leu Ser Glu Met Leu Met Asp Gly Gly Arg Gly
165 170 175
Asp Leu Asn Pro Ala Gly Ala Glu Tyr Gly Thr Gly Tyr Cys Asp Ala
180 185 190
Gln Cys Phe Lys Leu Asp Phe Ile Asn Gly Glu Ala Asn Ile Asp Gln
195 200 205
Lys His Gly Ala Cys Cys Asn Glu Met Asp Ile Phe Glu Ser Asn Ser
210 215 220
Arg Ala Lys Thr Phe Val Pro His Pro Cys Asn Ile Thr Gln Val Tyr
225 230 235 240
Lys Cys Glu Gly Glu Asp Glu Cys Gly Gln Pro Val Gly Val Cys Asp
245 250 255
Lys Trp Gly Cys Gly Phe Asn Glu Tyr Lys Trp Gly Val Glu Ser Phe
260 265 270
Tyr Gly Arg Gly Ser Gln Phe Ala Ile Asp Ser Ser Lys Lys Phe Thr
275 280 285
Val Thr Thr Gln Phe Leu Thr Asp Asn Gly Lys Glu Asp Gly Val Leu
290 295 300
Val Glu Ile Arg Arg Leu Trp His Gln Asp Gly Lys Leu Ile Lys Asn
305 310 315 320
Thr Ala Ile Gln Val Glu Glu Asn Tyr Ser Thr Asp Ser Val Ser Thr
325 330 335

```

Glu Phe Cys Glu Lys Thr Ala Ser Phe Thr Met Gln Arg Gly Gly Leu
 340 345 350
 Lys Ala Met Gly Glu Ala Ile Gly Arg Gly Met Val Leu Val Phe Ser
 355 360 365
 Ile Trp Ala Asp Asp Ser Gly Phe Met Asn Trp Leu Asp Ala Glu Gly
 370 375 380
 Asn Gly Pro Cys Ser Ala Thr Glu Gly Asp Pro Lys Glu Ile Val Lys
 385 390 395 400
 Asn Lys Pro Asp Ala Arg Val Thr Phe Ser Asn Ile Arg Ile Gly Glu
 405 410 415
 Val Gly Ser Thr Tyr Ala Pro Gly Gly Lys Cys Gly Val Lys Ser Arg
 420 425 430
 Val Ala Arg Gly Leu Thr Ala Ser
 435 440

<210> 21

<211> 1380

<212> DNA

<213> *Trichoderma reesei*

<400> 21

atggcgccct cagttacact gccgttgacc acggccatcc tggccattgc ccggtcgtc 60
 gccgcccagc aaccgggtac cagcaccccc gaggtccatc ccaagttgac aacctacaag 120
 tgtacaaagt ccgggggggtg cgtggcccag gacacctcgg tggtccttga ctggaactac 180
 cgctggatgc acgacgcaaa ctacaactcg tgcaccgtca acggcggcgt caacaccacg 240
 ctctgccectg acgagggcag ctgtggcaag aactgcttca tcgagggcgt cgactacgcc 300
 gcctcggggcg tcacgacctc gggcagcagc ctcaaccatga accagtacat gcccagcagc 360
 tctggcgggt acagcagcgt ctctcctcgg ctgtatctcc tggactctga cggtgagtac 420
 gtgatgctga agctcaacgg ccaggagctg agcttcgacg tcgacctctc tgctctgccg 480
 tgtggagaga acggctcgt ctacctgtct cagatggacg agaacggggg cgccaaccag 540
 tataacacgg ccggtgccaa ctacgggagc ggctactcgg atgctcagtg ccccgccag 600
 acatggagga acggcaccct caactactagc caccagggct tctgctgcaa cgagatggat 660
 atcctggagg gcaactcgag ggcgaatgcc ttgaccctc actcttgca ggcacggcc 720
 tgcgactctg ccggttgccg cttcaacccc tatggcagcg gctacaaaag ctactacggc 780
 cccggagata ccggttgacac ctccaagacc ttcaccatca tcaccagtt caacacggac 840
 aacggctcgc cctcgggcaa ccttgtgagc atcaccgcga agtaccagca aaacggcgtc 900
 gacatcccca ggcgccagcc cggcggcgac accatctcgt cctgcccgtc cgctcagcc 960
 tacggcggcc tcgccaccat gggcaaggcc ctgagcagcg gcatggtgct cgtgttcagc 1020
 atttggaacg acaacagcca gtacatgaac tggctcgaca gcggcaacgc gcgcccctgc 1080
 agcagcaccg agggcaaccc atccaacatc ctggccaaca accccaacac gcacgtcgtc 1140
 ttctccaaca tccgctgggg agacattggg tctactacga actcgactgc gccccgcc 1200
 ccgctcgtc ccagcagcagc gttttcgact acacggagga gctcgacgac ttcgagcagc 1260
 ccgagctgca cgcagactca ctggggggcag tgcggtggca ttgggtacag cgggtgcaag 1320
 acgtgcacgt cgggcactac gtgccagtat agcaacgact actactcgca atgccttag 1380

<210> 22

<211> 459

<212> PRT

<213> *Trichoderma reesei*

<400> 22

Met Ala Pro Ser Val Thr Leu Pro Leu Thr Thr Ala Ile Leu Ala Ile
 1 5 10 15
 Ala Arg Leu Val Ala Ala Gln Gln Pro Gly Thr Ser Thr Pro Glu Val
 20 25 30
 His Pro Lys Leu Thr Thr Tyr Lys Cys Thr Lys Ser Gly Gly Cys Val
 35 40 45
 Ala Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Met His
 50 55 60
 Asp Ala Asn Tyr Asn Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr
 65 70 75 80
 Leu Cys Pro Asp Glu Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly
 85 90 95
 Val Asp Tyr Ala Ala Ser Gly Val Thr Thr Ser Gly Ser Ser Leu Thr
 100 105 110
 Met Asn Gln Tyr Met Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser
 115 120 125

Pro Arg Leu Tyr Leu Leu Asp Ser Asp Gly Glu Tyr Val Met Leu Lys
 130 135 140
 Leu Asn Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro
 145 150 155 160
 Cys Gly Glu Asn Gly Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly
 165 170 175
 Gly Ala Asn Gln Tyr Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr
 180 185 190
 Cys Asp Ala Gln Cys Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn
 195 200 205
 Thr Ser His Gln Gly Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly
 210 215 220
 Asn Ser Arg Ala Asn Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala
 225 230 235 240
 Cys Asp Ser Ala Gly Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Lys
 245 250 255
 Ser Tyr Tyr Gly Pro Gly Asp Thr Val Asp Thr Ser Lys Thr Phe Thr
 260 265 270
 Ile Ile Thr Gln Phe Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu
 275 280 285
 Val Ser Ile Thr Arg Lys Tyr Gln Gln Asn Gly Val Asp Ile Pro Ser
 290 295 300
 Ala Gln Pro Gly Gly Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala
 305 310 315 320
 Tyr Gly Gly Leu Ala Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val
 325 330 335
 Leu Val Phe Ser Ile Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu
 340 345 350
 Asp Ser Gly Asn Ala Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser
 355 360 365
 Asn Ile Leu Ala Asn Asn Pro Asn Thr His Val Val Phe Ser Asn Ile
 370 375 380
 Arg Trp Gly Asp Ile Gly Ser Thr Thr Asn Ser Thr Ala Pro Pro Pro
 385 390 395 400
 Pro Pro Ala Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr
 405 410 415
 Thr Ser Ser Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly
 420 425 430
 Gly Ile Gly Tyr Ser Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys
 435 440 445
 Gln Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu
 450 455

<210> 23

<211> 2586

<212> DNA

<213> *Aspergillus oryzae*

<400> 23

```

atgaagcttg gttggatcga ggtggcgcga ttggcggctg cctcagtagt cagtgccaaag      60
gatgatctcg cgtactcccc tcttttctac ccttccccat gggcagatgg tcagggtgaa      120
tgggcggaag tatacaaacg cgctgtagac atagtttccc agatgacgtt gacagagaaa      180
gtcaacttaa cgactggaac aggatggcaa ctagagaggt gtggttgaca aactggcagt      240
gttcccagac tcaacatccc cagcttgtgt ttgcaggata gtcctcttgg tattcgtttc      300
tcggactaca attcagcttt ccctgctgggt gttaatgtcg ctgccacctg ggacaagacg      360
ctcgcctacc ttctgtggtca ggcaatgggt gaggagtcca gtgataaggg tattgacgtt      420
cagctgggtc ctgctgctgg ccctctcggg gctcatccgg atggcggtag aaactgggaa      480
ggtttctcac cagatccagc cctcaccggt gtactttttg cggagacgat taagggtatt      540
caagatgctg gtgtcattgc gacagctaag cattatatca tgaacgaaca agagcatttc      600
cgccaacaac ccgaggctgc gggttacgga ttcaacgtaa gcgacagttt gagttccaac      660
gttgatgaca agactatgca tgaattgtac ctctggccct tcgcggatgc agtacgcgct      720
ggagtcggtg ctgtcatgtg ctcttacaac caaatcaaca acagctacgg ttgcgagaat      780
agcgaaactc tgaacaagct tttgaaggcg gagcttggtt tccaaggctt cgtcatgagt      840
gattggaccg ctcatcacag cggcgtaggc gctgctttag caggtctgga tatgtcgatg      900
cccggatgatg ttaccttoga tagtggtacg tctttctggg gtgcaaactt gacggtcggg      960
  
```

gtcottaacg gtacaatccc ccaatggcgt gttgatgaca tggctgtccg tatcatggcc 1020
 gcttattaca aggttggccg cgacaccaa tacaccctc ccaacttcag ctctgtggacc 1080
 agggacgaat atggtttcgc gcataaccat gtttcggaag gtgcttacga gagggtcaac 1140
 gaattcgtgg acgtgcaacg cgatcatgcc gacctaattc gtcgcatcgg cgcgcagagc 1200
 actgttctgc tgaagaacaa ggggtgccttg cccttgagcc gcaaggaaaa gctggtcgcc 1260
 cttctgggag aggatgcggg ttccaactcg tggggcgcta acggctgtga tgaccgtggt 1320
 tgcgataacg gtacccttgc catggcctgg ggtagcggta ctgcgaattt cccatacctc 1380
 gtgacaccag agcaggcgat tcagaacgaa gttcttcagg gccgtggtaa tgccttcgcc 1440
 gtgaccgaca gttgggcgct cgacaagatc gctgcggctg cccgccaggc cagcgtatct 1500
 ctctgtttcg tcaactccga ctcaggagaa ggctatctta gtgtggatgg aaatgagggc 1560
 gatcgtaaca acatcactct gtggaagaac ggcgacaatg tggtaaacgac cgcagcgaat 1620
 aactgtaaca acaccgttgt catcatccac tccgtcggac cagttttgat cgatgaatgg 1680
 tatgaccacc ccaatgtcac tggattctc tgggctggtc tgccaggcca ggagtctggt 1740
 aactccattg ccgatgtgct gtacggctgt gtcaaccctg gcgccaagtc tcctttcaact 1800
 tggggcaaga cccgggagtc gtatggttct cccttgggtca aggatgcca caatggcaac 1860
 ggagcgcccc agtctgattt caccagggtt gttttcatcg attaccgcca tttcgataag 1920
 ttcaatgaga cccctatcta cgagtttggc tacggcttga gctacaccac cttcgagctc 1980
 tccgacctoc atgttcagcc cctgaacgcy tcccgatata ctcccaccag tggcatgact 2040
 gaagctgcaa agaactttgg tgaatttggc gatgcgtcgg agtacgtgta tccggagggg 2100
 ctggaaagga tccatgagtt tatctatccc tggatcaact ctaccgacct gaaggcatcg 2160
 tctgacgatt ctaactacgg ctgggaagac tccaagtata ttcccgaagg cgccacggat 2220
 gggctctgcc agccccgttt gcccgctagt ggtggtgccg gaggaaacc cggctctgtac 2280
 gaggatcttt tccgcgtctc tgtgaaggtc aagaacacgg gcaatgtcgc cggatgatgaa 2340
 gttcctcagc tgtacgtttc cctaggcggc ccgaatgagc ccaagggtgg actgcgcaag 2400
 tttgagcgta ttcacttggc cccttcgcag gaggccgtgt ggacaacgac ccttaccctg 2460
 cgtgaccttg caaactggga cgtttcggct caggactgga ccgtcactcc ttaccccaag 2520
 acgatctacg ttggaaactc ctcacggaaa ctgcccgtcc aggcctcgtc gcctaaggcc 2580
 cagtaa 2586

<210> 24

<211> 861

<212> PRT

<213> *Aspergillus oryzae*

<400> 24

Met Lys Leu Gly Trp Ile Glu Val Ala Ala Leu Ala Ala Ala Ser Val
 1 5 10 15
 Val Ser Ala Lys Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser
 20 25 30
 Pro Trp Ala Asp Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala
 35 40 45
 Val Asp Ile Val Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr
 50 55 60
 Thr Gly Thr Gly Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser
 65 70 75 80
 Val Pro Arg Leu Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu
 85 90 95
 Gly Ile Arg Phe Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn
 100 105 110
 Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Gln Ala
 115 120 125
 Met Gly Glu Glu Phe Ser Asp Lys Gly Ile Asp Val Gln Leu Gly Pro
 130 135 140
 Ala Ala Gly Pro Leu Gly Ala His Pro Asp Gly Gly Arg Asn Trp Glu
 145 150 155 160
 Gly Phe Ser Pro Asp Pro Ala Leu Thr Gly Val Leu Phe Ala Glu Thr
 165 170 175
 Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
 180 185 190
 Ile Met Asn Glu Gln Glu His Phe Arg Gln Gln Pro Glu Ala Ala Gly
 195 200 205
 Tyr Gly Phe Asn Val Ser Asp Ser Leu Ser Ser Asn Val Asp Asp Lys
 210 215 220
 Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
 225 230 235 240

Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
 245 250 255
 Gly Cys Glu Asn Ser Glu Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
 260 265 270
 Gly Phe Gln Gly Phe Val Met Ser Asp Trp Thr Ala His His Ser Gly
 275 280 285
 Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Val
 290 295 300
 Thr Phe Asp Ser Gly Thr Ser Phe Trp Gly Ala Asn Leu Thr Val Gly
 305 310 315 320
 Val Leu Asn Gly Thr Ile Pro Gln Trp Arg Val Asp Asp Met Ala Val
 325 330 335
 Arg Ile Met Ala Ala Tyr Tyr Lys Val Gly Arg Asp Thr Lys Tyr Thr
 340 345 350
 Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Phe Ala His
 355 360 365
 Asn His Val Ser Glu Gly Ala Tyr Glu Arg Val Asn Glu Phe Val Asp
 370 375 380
 Val Gln Arg Asp His Ala Asp Leu Ile Arg Arg Ile Gly Ala Gln Ser
 385 390 395 400
 Thr Val Leu Leu Lys Asn Lys Gly Ala Leu Pro Leu Ser Arg Lys Glu
 405 410 415
 Lys Leu Val Ala Leu Leu Gly Glu Asp Ala Gly Ser Asn Ser Trp Gly
 420 425 430
 Ala Asn Gly Cys Asp Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
 435 440 445
 Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu
 450 455 460
 Gln Ala Ile Gln Asn Glu Val Leu Gln Gly Arg Gly Asn Val Phe Ala
 465 470 475 480
 Val Thr Asp Ser Trp Ala Leu Asp Lys Ile Ala Ala Ala Ala Arg Gln
 485 490 495
 Ala Ser Val Ser Leu Val Phe Val Asn Ser Asp Ser Gly Glu Gly Tyr
 500 505 510
 Leu Ser Val Asp Gly Asn Glu Gly Asp Arg Asn Asn Ile Thr Leu Trp
 515 520 525
 Lys Asn Gly Asp Asn Val Val Lys Thr Ala Ala Asn Asn Cys Asn Asn
 530 535 540
 Thr Val Val Ile Ile His Ser Val Gly Pro Val Leu Ile Asp Glu Trp
 545 550 555 560
 Tyr Asp His Pro Asn Val Thr Gly Ile Leu Trp Ala Gly Leu Pro Gly
 565 570 575
 Gln Glu Ser Gly Asn Ser Ile Ala Asp Val Leu Tyr Gly Arg Val Asn
 580 585 590
 Pro Gly Ala Lys Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
 595 600 605
 Gly Ser Pro Leu Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln
 610 615 620
 Ser Asp Phe Thr Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys
 625 630 635 640
 Phe Asn Glu Thr Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr
 645 650 655
 Thr Phe Glu Leu Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg
 660 665 670
 Tyr Thr Pro Thr Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu
 675 680 685
 Ile Gly Asp Ala Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile
 690 695 700
 His Glu Phe Ile Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser
 705 710 715 720
 Ser Asp Asp Ser Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu
 725 730 735
 Gly Ala Thr Asp Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly

			740						745				750			
Ala	Gly	Gly	Asn	Pro	Gly	Leu	Tyr	Glu	Asp	Leu	Phe	Arg	Val	Ser	Val	
			755				760					765				
Lys	Val	Lys	Asn	Thr	Gly	Asn	Val	Ala	Gly	Asp	Glu	Val	Pro	Gln	Leu	
			770				775					780				
Tyr	Val	Ser	Leu	Gly	Gly	Pro	Asn	Glu	Pro	Lys	Val	Val	Leu	Arg	Lys	
			785				790				795				800	
Phe	Glu	Arg	Ile	His	Leu	Ala	Pro	Ser	Gln	Glu	Ala	Val	Trp	Thr	Thr	
				805					810					815		
Thr	Leu	Thr	Arg	Arg	Asp	Leu	Ala	Asn	Trp	Asp	Val	Ser	Ala	Gln	Asp	
			820					825					830			
Trp	Thr	Val	Thr	Pro	Tyr	Pro	Lys	Thr	Ile	Tyr	Val	Gly	Asn	Ser	Ser	
			835				840					845				
Arg	Lys	Leu	Pro	Leu	Gln	Ala	Ser	Leu	Pro	Lys	Ala	Gln				
			850				855					860				

<210> 25

<211> 2586

<212> DNA

<213> *Aspergillus oryzae*

<400> 25

atgaagcttg	ggtggatcga	ggtggccgca	ttggcgctg	cctcagtagt	cagtgccaaag	60
gatgatctcg	cgtactcccc	tcctttctac	ccttccccat	gggcagatgg	tcagggtgaa	120
tgggcggaag	tatacaaacg	cgctgtagac	atagtttccc	agatgacggt	gacagagaaa	180
gtcaacttaa	cgactggaac	aggatggcaa	ctagagaggt	gtgttggaca	aactggcagt	240
gttcccagac	tcaacatccc	cagcttgtgt	ttgcaggata	gtcctcttgg	tattcgtttc	300
tcggactaca	attcagcttt	ccctgcccgt	gttaatgtcg	ctgccacctg	ggacaagacg	360
ctgcctacc	ttcgtgggtca	ggcaatgggt	gaggagtcca	gtgataagg	tattgacgtt	420
cagctgggtc	ctgctgctgg	ccctctcggt	gtcatccgg	atggcggtag	aaactgggaa	480
agtttctcac	cagatccagc	cctcaccggt	gtactttttg	cggagacgat	taagggtatt	540
caagatgctg	gtgtcattgc	gacagctaag	cattatatca	tgaacgaaca	agagcatttc	600
cgccaacaac	ccgaggctgc	gggttacgga	ttcaacgtaa	gcgacagttt	gagttccaac	660
gttgatgaca	agactatgca	tgaattgtac	ctctggccct	tcgcggatgc	agtacgcgct	720
ggagtcgggtg	ctgttatgtg	ctcttacaac	caaatcaaca	acagctacgg	ttgcgagaat	780
agcgaaactc	tgaacaagct	tttgaaggcg	gagcttgggt	tccaaggctt	cgatcatgag	840
gattggaccg	ctcaacacag	cggcgtaggc	gctgctttag	caggtctgga	tatgtcgatg	900
cccggtagtg	ttaccttcga	tagtggtagc	tcttctggg	gtgcaaactt	gacggtcgg	960
gtccttaacg	gtacaatccc	ccaatggcgt	gttgatgaca	tggctgtccg	tatcatggcc	1020
gcttattaca	aggttggccg	cgacaccaa	tacaccctc	ccaacttcag	ctcgtggacc	1080
agggacgaat	atggtttcgc	gcataaccat	gtttcgggaag	gtgcttacga	gagggtcaac	1140
gaattcgtgg	acgtgcaacg	cgatcatgcc	gacctaatcc	gtcgcacg	cgcgacagac	1200
actgttctgc	tgaagaacaa	gggtgccttg	cccttgagcc	gcaaggaaaa	gctggctgcc	1260
cttctgggag	aggatgcggg	ttccaactcg	tggggcgcta	acggctgtga	tgaccgtggt	1320
tgcgataacg	gtacccttgc	catggcctgg	ggtagcggta	ctgcgaattt	cccataacct	1380
gtgacaccag	agcaggcgat	tcagaacgaa	gttcttcag	gocgtggtaa	tgtcttcgcc	1440
gtgaccgaca	gttggcgct	cgacaagatc	gctgcggctg	cccgccaggc	cagcgtatct	1500
ctcgtgttcg	tcaactccga	ctcaggagaa	gtctatctta	gtgtggatgg	aaactgaggc	1560
gatcgtaaca	acatcactct	gtggaagaac	ggcgacaatg	tggtcaagac	cgacgcgaat	1620
aactgtaaca	acaccgttgt	catcatccac	tccgtcggac	cagttttgat	cgatgaatgg	1680
tatgaccacc	ccaatgtcac	tggtattctc	tgggctggtc	tgccaggcca	ggagtctggt	1740
aactccattg	ccgatgtgct	gtacggctcg	gtcaaccctg	gcgccaagtc	tcctttcact	1800
tggggcaaga	cccgggagtc	gtatggttct	cccttgggtca	aggatgcca	caatggcaac	1860
ggagcgcccc	agtctgattt	caccaggggt	gttttcatcg	attaccgcca	tttcgataag	1920
ttcaatgaga	cccctatcta	cgagtttggc	tacggcttga	gctacaccac	cttcgagctc	1980
tccgacctcc	atgttcagcc	ctggaacg	tcccgatata	ctcccaccag	tggcagatct	2040
gaagctgcaa	agaactttgg	tgaattggc	gatgcctcgg	agtacgtgta	tccggagggg	2100
ctggaaagga	tccatgagtt	tatctatccc	tggatcaact	ctaccgacct	gaaggcatcg	2160
tctgacgatt	ctaactacgg	ctgggaagac	tccaagtata	ttcccgaagg	cgccacggat	2220
gggtctgccc	agccccgttt	gcccgtag	ggtggtgccc	gaggaaaccc	cggtctgtac	2280
gaggatcttt	tccgcgtctc	tgtgaaggtc	aagaacacgg	gcaatgtcgc	cggtgatgaa	2340
gttccctcagc	tgtacgtttc	cctaggcggc	ccgaatgagc	ccaaggtggt	actgcgcaag	2400
tttgagcgta	ttcacttggc	cccttcgag	gaggccgtgt	ggacaacgac	ccttaccctg	2460
cgtgaccttg	caaactggga	cgtttcggct	caggactgga	ccgtcactcc	ttaccccaag	2520
acgatctacg	ttggaaactc	ctcacgga	ctgccgctcc	aggcctcgtc	gcctaaggcc	2580

cagtaa
 <210> 26
 <211> 861
 <212> PRT
 <213> *Aspergillus oryzae*
 <400> 26
 Met Lys Leu Gly Trp Ile Glu Val Ala Ala Leu Ala Ala Ala Ser Val
 1 5 10 15
 Val Ser Ala Lys Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser
 20 25 30
 Pro Trp Ala Asp Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala
 35 40 45
 Val Asp Ile Val Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr
 50 55 60
 Thr Gly Thr Gly Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser
 65 70 75 80
 Val Pro Arg Leu Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu
 85 90 95
 Gly Ile Arg Phe Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn
 100 105 110
 Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Gln Ala
 115 120 125
 Met Gly Glu Glu Phe Ser Asp Lys Gly Ile Asp Val Gln Leu Gly Pro
 130 135 140
 Ala Ala Gly Pro Leu Gly Ala His Pro Asp Gly Gly Arg Asn Trp Glu
 145 150 155 160
 Ser Phe Ser Pro Asp Pro Ala Leu Thr Gly Val Leu Phe Ala Glu Thr
 165 170 175
 Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
 180 185 190
 Ile Met Asn Glu Gln Glu His Phe Arg Gln Gln Pro Glu Ala Ala Gly
 195 200 205
 Tyr Gly Phe Asn Val Ser Asp Ser Leu Ser Ser Asn Val Asp Asp Lys
 210 215 220
 Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
 225 230 235 240
 Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
 245 250 255
 Gly Cys Glu Asn Ser Glu Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
 260 265 270
 Gly Phe Gln Gly Phe Val Met Ser Asp Trp Thr Ala Gln His Ser Gly
 275 280 285
 Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Val
 290 295 300
 Thr Phe Asp Ser Gly Thr Ser Phe Trp Gly Ala Asn Leu Thr Val Gly
 305 310 315 320
 Val Leu Asn Gly Thr Ile Pro Gln Trp Arg Val Asp Asp Met Ala Val
 325 330 335
 Arg Ile Met Ala Ala Tyr Tyr Lys Val Gly Arg Asp Thr Lys Tyr Thr
 340 345 350
 Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Phe Ala His
 355 360 365
 Asn His Val Ser Glu Gly Ala Tyr Glu Arg Val Asn Glu Phe Val Asp
 370 375 380
 Val Gln Arg Asp His Ala Asp Leu Ile Arg Arg Ile Gly Ala Gln Ser
 385 390 395 400
 Thr Val Leu Leu Lys Asn Lys Gly Ala Leu Pro Leu Ser Arg Lys Glu
 405 410 415
 Lys Leu Val Ala Leu Leu Gly Glu Asp Ala Gly Ser Asn Ser Trp Gly
 420 425 430
 Ala Asn Gly Cys Asp Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
 435 440 445
 Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu

450						455						460					
Gln	Ala	Ile	Gln	Asn	Glu	Val	Leu	Gln	Gly	Arg	Gly	Asn	Val	Phe	Ala		
465					470					475					480		
Val	Thr	Asp	Ser	Trp	Ala	Leu	Asp	Lys	Ile	Ala	Ala	Ala	Ala	Arg	Gln		
			485					490						495			
Ala	Ser	Val	Ser	Leu	Val	Phe	Val	Asn	Ser	Asp	Ser	Gly	Glu	Gly	Tyr		
			500					505					510				
Leu	Ser	Val	Asp	Gly	Asn	Glu	Gly	Asp	Arg	Asn	Asn	Ile	Thr	Leu	Trp		
		515					520					525					
Lys	Asn	Gly	Asp	Asn	Val	Val	Lys	Thr	Ala	Ala	Asn	Asn	Cys	Asn	Asn		
	530					535						540					
Thr	Val	Val	Ile	Ile	His	Ser	Val	Gly	Pro	Val	Leu	Ile	Asp	Glu	Trp		
545					550					555				560			
Tyr	Asp	His	Pro	Asn	Val	Thr	Gly	Ile	Leu	Trp	Ala	Gly	Leu	Pro	Gly		
				565				570						575			
Gln	Glu	Ser	Gly	Asn	Ser	Ile	Ala	Asp	Val	Leu	Tyr	Gly	Arg	Val	Asn		
			580					585					590				
Pro	Gly	Ala	Lys	Ser	Pro	Phe	Thr	Trp	Gly	Lys	Thr	Arg	Glu	Ser	Tyr		
		595					600					605					
Gly	Ser	Pro	Leu	Val	Lys	Asp	Ala	Asn	Asn	Gly	Asn	Gly	Ala	Pro	Gln		
	610					615						620					
Ser	Asp	Phe	Thr	Gln	Gly	Val	Phe	Ile	Asp	Tyr	Arg	His	Phe	Asp	Lys		
625					630					635				640			
Phe	Asn	Glu	Thr	Pro	Ile	Tyr	Glu	Phe	Gly	Tyr	Gly	Leu	Ser	Tyr	Thr		
				645					650					655			
Thr	Phe	Glu	Leu	Ser	Asp	Leu	His	Val	Gln	Pro	Leu	Asn	Ala	Ser	Arg		
			660					665					670				
Tyr	Thr	Pro	Thr	Ser	Gly	Met	Thr	Glu	Ala	Ala	Lys	Asn	Phe	Gly	Glu		
		675					680						685				
Ile	Gly	Asp	Ala	Ser	Glu	Tyr	Val	Tyr	Pro	Glu	Gly	Leu	Glu	Arg	Ile		
	690					695					700						
His	Glu	Phe	Ile	Tyr	Pro	Trp	Ile	Asn	Ser	Thr	Asp	Leu	Lys	Ala	Ser		
705					710					715				720			
Ser	Asp	Asp	Ser	Asn	Tyr	Gly	Trp	Glu	Asp	Ser	Lys	Tyr	Ile	Pro	Glu		
				725						730				735			
Gly	Ala	Thr	Asp	Gly	Ser	Ala	Gln	Pro	Arg	Leu	Pro	Ala	Ser	Gly	Gly		
			740					745					750				
Ala	Gly	Gly	Asn	Pro	Gly	Leu	Tyr	Glu	Asp	Leu	Phe	Arg	Val	Ser	Val		
		755					760					765					
Lys	Val	Lys	Asn	Thr	Gly	Asn	Val	Ala	Gly	Asp	Glu	Val	Pro	Gln	Leu		
	770					775					780						
Tyr	Val	Ser	Leu	Gly	Gly	Pro	Asn	Glu	Pro	Lys	Val	Val	Leu	Arg	Lys		
785					790					795				800			
Phe	Glu	Arg	Ile	His	Leu	Ala	Pro	Ser	Gln	Glu	Ala	Val	Trp	Thr	Thr		
				805					810					815			
Thr	Leu	Thr	Arg	Arg	Asp	Leu	Ala	Asn	Trp	Asp	Val	Ser	Ala	Gln	Asp		
			820					825					830				
Trp	Thr	Val	Thr	Pro	Tyr	Pro	Lys	Thr	Ile	Tyr	Val	Gly	Asn	Ser	Ser		
		835					840					845					
Arg	Lys	Leu	Pro	Leu	Gln	Ala	Ser	Leu	Pro	Lys	Ala	Gln					
	850					855						860					

<210> 27

<211> 3060

<212> DNA

<213> *Aspergillus fumigatus*

<400> 27

atgagatttcg	ggtggctcga	ggtggccgct	ctgacggccg	cttctgtagc	caatgccag	60
gtttgtgatg	ctttcccgtc	attgtttcgg	atatagttag	caatagtcac	ggaaataatc	120
aggaattggc	tttctctcca	ccattctacc	cttcgccttg	ggctgatggc	cagggagagt	180
gggcagatgc	ccatcgacgc	gccgtcgaga	tcgtttctca	gatgacactg	gctggagaag	240
ttaaccttac	aacgggtact	gggtgggttg	cgactttttt	gttgacagtg	agctttcttc	300
actgaccatc	tacacagatg	ggaaatggac	cgatgcgtcg	gtcaaaccgg	cagcgttccc	360
aggtaaagctt	gcaattctgc	aacaacgtgc	aagtgtagtt	gctaaaacgc	ggtggtgacg	420

acttgggtatc aactggggtc tttgtggcca ggattcccct ttgggtatcc gtttctgtga 480
 gctataccccg cggagtcttt cagtccttgt attatgtgct gatgattgtc tctgtatagc 540
 tgacctcaac tccgccttcc ctgctggtag taatgtcgcc gcgacatggg acaagacact 600
 cgcctacctt cgtggcaagg ccatgggtga ggaattcaac gacaagggcg tggacatttt 660
 gctggggcct gctgctggtc ctctcggcaa ataccgggac ggcggcagaa tctgggaagg 720
 cttctctcct gatccgggtc tcaactgggt acttttcgcc gaaactatca agggatcca 780
 agacgcgggt gtgattgcta ctgccaagca ttacattctg aatgaacagg agcatttccg 840
 acaggttggc gaggcccagg gatatggta caacatcacg gagacgatca gctccaacgt 900
 ggatgacaag accatgcacg agttgtacct ttgggtgagta gttgacactg caaatgagga 960
 ccttgattga tttgactgac ctggaatgca ggcccttgc agatgctgtg cgcggtaaga 1020
 ttttccgtag acttgacctc gcgacgaaga aatcgctgac gaaccatcgt agctggcggt 1080
 ggcgctgtca tgtgttcta caatcaaadc aacaacagct acggttgtca aaacagtcaa 1140
 actctcaaca agctcctcaa ggctgagctg ggcttccaag gcttcgcatc gatgactgg 1200
 agcgtcacc acagcgggtg cggcgtgccc ctgctgggtg tggatatgtc gatgcctgga 1260
 gacatttctt tgcagcagcg actctccttc tggggcacga acctaaactgt cagtgttctt 1320
 aacggcaccg ttccagcctg gcgtgtcgat gacatggctg ttcgtatcat gaccgcgtac 1380
 tacaaggttg gtcgtgaccg tcttcgtatt cccctaaact tcagctcctg gaccgggat 1440
 gagtacggct gggagcattc tgetgtctcc gagggagcct ggaccaaggc gaacgacttc 1500
 gtcaatgtgc agcgcagtca ctctcagatc atccgtgaga ttgggtgccg tagtacagtg 1560
 ctcttgaaga acacgggtgc tcttccttgc accggcaagg aggttaaagt ggggttctc 1620
 ggtgaagacg ctgggttccaa cccgtggggg gctaacggct gccccgaccg cggetgtgat 1680
 aacggcactc ttgctatggc ctggggtagt ggtactgcca acttccctta ccttgcacc 1740
 cccgagcagg ctatccagcg agaggtcatc agcaacggcg gcaatgtctt tgctgtgact 1800
 gataacgggg ctctcagcca gatggcagat gttgcatctc aatccagggt agtgcgggct 1860
 cttagaaaaa gaacgttctc tgaatgaagt tttttaacca ttgcaaacag cgtgtctttg 1920
 gtgtttgtca acgccgactc tggagagggg ttcatcagtg tcgacggcaa cgagggtagc 1980
 cgcaaaaatc tcaactgtg gaagaacggc gagggcgtca ttgacactgt tgtcagccac 2040
 tgcaacaaca cgatttgggt tattcacagt gttgggcccg tcttgatcga ccggtggtat 2100
 gataacccca acgtcactgc catcatctgg gccggcttgc ccggtcagga gagtggcaac 2160
 tccctggctc acgtgctcta tggccgcgct aaccccagcg ccaagacccc gttcacctgg 2220
 ggcaagactc gggagtctta cggggctccc ttgctcaccg agcctaaca tggcaatggt 2280
 gtcccccagg atgatttcaa cgagggcgtc ttcattgact accgtcactt tgacaagcgc 2340
 aatgagacc ccatthtga gtttggccat ggcttgagct acaccaactt tggttactct 2400
 caccttcggg ttccagccct caatagttcg agttcggcat atgtcccagc tagcggagag 2460
 accaagcctg cgccaaccta tggtagatc ggtagtgcg ccgactacct gtatcccag 2520
 ggtctcaaaa gaattacaa gtttatttac ccttggctca actcgaccga cctcgaggat 2580
 tcttctgacg acccgaacta cggctgggag gactcggagt acattcccga aggcgctagg 2640
 gatgggtctc ctcaaccct cctgaaggct ggcggcgtc ctgggtggtaa ccctaccctt 2700
 tatcaggatc ttgttagggt gtcggccacc ataaccaaca ctggtaacgt cgccggttat 2760
 gaagtcctc aattggtgag tgaccgcgat gttccttgcg ttgcaatttg gctaactcgc 2820
 ttctagtagt tttcactggg cggaccgaac gagcctcggg tcgttctgcg caagttcgac 2880
 cgaatcttcc tggctcctgg ggagcaaaa gtttggacca cgactcttaa ccgctcgtgat 2940
 ctgcaccaat gggatgtgga ggctcaggac tgggtcatca caaagtacc caagaaagt 3000
 cacgtcggca gctcctcgcg taagctgct ctgagagcgc ctctgccccg tgtctactag 3060

<210> 28

<211> 863

<212> PRT

<213> *Aspergillus fumigatus*

<400> 28

Met Arg Phe Gly Trp Leu Glu Val Ala Ala Leu Thr Ala Ala Ser Val
 1 5 10 15
 Ala Asn Ala Gln Glu Leu Ala Phe Ser Pro Pro Phe Tyr Pro Ser Pro
 20 25 30
 Trp Ala Asp Gly Gln Gly Glu Trp Ala Asp Ala His Arg Arg Ala Val
 35 40 45
 Glu Ile Val Ser Gln Met Thr Leu Ala Glu Lys Val Asn Leu Thr Thr
 50 55 60
 Gly Thr Gly Trp Glu Met Asp Arg Cys Val Gly Gln Thr Gly Ser Val
 65 70 75 80
 Pro Arg Leu Gly Ile Asn Trp Gly Leu Cys Gly Gln Asp Ser Pro Leu
 85 90 95
 Gly Ile Arg Phe Ser Asp Leu Asn Ser Ala Phe Pro Ala Gly Thr Asn
 100 105 110

Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Lys Ala
 115 120 125
 Met Gly Glu Glu Phe Asn Asp Lys Gly Val Asp Ile Leu Leu Gly Pro
 130 135 140
 Ala Ala Gly Pro Leu Gly Lys Tyr Pro Asp Gly Gly Arg Ile Trp Glu
 145 150 155 160
 Gly Phe Ser Pro Asp Pro Val Leu Thr Gly Val Leu Phe Ala Glu Thr
 165 170 175
 Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
 180 185 190
 Ile Leu Asn Glu Gln Glu His Phe Arg Gln Val Gly Glu Ala Gln Gly
 195 200 205
 Tyr Gly Tyr Asn Ile Thr Glu Thr Ile Ser Ser Asn Val Asp Asp Lys
 210 215 220
 Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
 225 230 235 240
 Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
 245 250 255
 Gly Cys Gln Asn Ser Gln Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
 260 265 270
 Gly Phe Gln Gly Phe Val Met Ser Asp Trp Ser Ala His His Ser Gly
 275 280 285
 Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Ile
 290 295 300
 Ser Phe Asp Asp Gly Leu Ser Phe Trp Gly Thr Asn Leu Thr Val Ser
 305 310 315 320
 Val Leu Asn Gly Thr Val Pro Ala Trp Arg Val Asp Asp Met Ala Val
 325 330 335
 Arg Ile Met Thr Ala Tyr Tyr Lys Val Gly Arg Asp Arg Leu Arg Ile
 340 345 350
 Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Trp Glu His
 355 360 365
 Ser Ala Val Ser Glu Gly Ala Trp Thr Lys Val Asn Asp Phe Val Asn
 370 375 380
 Val Gln Arg Ser His Ser Gln Ile Ile Arg Glu Ile Gly Ala Ala Ser
 385 390 395 400
 Thr Val Leu Leu Lys Asn Thr Gly Ala Leu Pro Leu Thr Gly Lys Glu
 405 410 415
 Val Lys Val Gly Val Leu Gly Glu Asp Ala Gly Ser Asn Pro Trp Gly
 420 425 430
 Ala Asn Gly Cys Pro Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
 435 440 445
 Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu
 450 455 460
 Gln Ala Ile Gln Arg Glu Val Ile Ser Asn Gly Gly Asn Val Phe Ala
 465 470 475 480
 Val Thr Asp Asn Gly Ala Leu Ser Gln Met Ala Asp Val Ala Ser Gln
 485 490 495
 Ser Ser Val Ser Leu Val Phe Val Asn Ala Asp Ser Gly Glu Gly Phe
 500 505 510
 Ile Ser Val Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp
 515 520 525
 Lys Asn Gly Glu Ala Val Ile Asp Thr Val Val Ser His Cys Asn Asn
 530 535 540
 Thr Ile Val Val Ile His Ser Val Gly Pro Val Leu Ile Asp Arg Trp
 545 550 555 560
 Tyr Asp Asn Pro Asn Val Thr Ala Ile Ile Trp Ala Gly Leu Pro Gly
 565 570 575
 Gln Glu Ser Gly Asn Ser Leu Val Asp Val Leu Tyr Gly Arg Val Asn
 580 585 590
 Pro Ser Ala Lys Thr Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
 595 600 605
 Gly Ala Pro Leu Leu Thr Glu Pro Asn Asn Gly Asn Gly Ala Pro Gln

actatgacga caatgctatc ttgtcgttg tctcacagcc tgatgcaacc tgtatcgttt 1680
 ttgcaaatgc cgattccggg gaaggctaca tcaactgctga caacaactgg ggtgaccgca 1740
 acaatctgac cctctggcaa aatgccgatc aagtgattag cactgtcagc tcgcatgca 1800
 acaacacaat cgttgttctc cactctgtcg gaccagtgtt gctaaatggg atatatgagc 1860
 acccgaacat cacagctatt gtctgggcag ggatgccagg cgaagaatct ggcaatgctc 1920
 tcgtggatat tctttggggc aatgttaacc ctgccggtcg cactccgttc acctggggcca 1980
 aaagtcgaga ggactatggc actgatataa tgtaecagcc caacaacggc cagcgtgcgc 2040
 ctcagcagga tttcaccgag agcatctacc tgcactaccg ccatttcgac aaagctggta 2100
 tcgagccaat ttacgagttt ggattcggcc tctcctatac caccttcgaa tactctgacc 2160
 tccgtgttgt gaagaagtat gttcaacat acagtccac gaccggcacc ggtgctcaag 2220
 caccttccat cggacagcca cctagccaga acctggatac ctacaagttc cctgctacat 2280
 acaagtacat caaaaccttc atttatccct acctgaacag cactgtctcc ctccgcgctg 2340
 cttccaagga tcccgaatac ggtcgtacag actttatccc accccacgcg cgtgatggct 2400
 cccctcaacc tctcaacccc gctggagacc cagtggccag tggtggaac aacatgctct 2460
 acgacgaact ttacgaggtc actgcacaga tcaaaaacac tggcgacgtg gccggcgacg 2520
 aagtcgtcca gctttacgta gatctcgggg gtgacaaccc gcctcgtcag ttgagaaact 2580
 ttgacaggtt ttatctgctg cccggtcaga gctcaacatt ccgggctaca ttgacgcgcc 2640
 gtgatttgag caactgggat attgaggcgc agaactggcg agttacggaa tcgcctaaga 2700
 gagtgtatgt tggacggctg agtcgggatt tgccgctgag ctcaaatg gagtaatgat 2760
 catgtctacc aatagatggt gaatgtctgg tgtggatatt 2800
 <210> 30
 <211> 878
 <212> PRT
 <213> Penicillium brasilianum
 <400> 30
 Met Gln Gly Ser Thr Ile Phe Leu Ala Phe Ala Ser Trp Ala Ser Gln
 1 5 10 15
 Val Ala Ala Ile Ala Gln Pro Ile Gln Lys His Glu Pro Gly Phe Leu
 20 25 30
 His Gly Pro Gln Ala Ile Glu Ser Phe Ser Glu Pro Phe Tyr Pro Ser
 35 40 45
 Pro Trp Met Asn Pro His Ala Glu Gly Trp Glu Ala Ala Tyr Gln Lys
 50 55 60
 Ala Gln Asp Phe Val Ser Gln Leu Thr Ile Leu Glu Lys Ile Asn Leu
 65 70 75 80
 Thr Thr Gly Val Gly Trp Glu Asn Gly Pro Cys Val Gly Asn Thr Gly
 85 90 95
 Ser Ile Pro Arg Leu Gly Phe Lys Gly Phe Cys Thr Gln Asp Ser Pro
 100 105 110
 Gln Gly Val Arg Phe Ala Asp Tyr Ser Ser Ala Phe Thr Ser Ser Gln
 115 120 125
 Met Ala Ala Ala Thr Phe Asp Arg Ser Ile Leu Tyr Gln Arg Gly Gln
 130 135 140
 Ala Met Ala Gln Glu His Lys Ala Lys Gly Ile Thr Ile Gln Leu Gly
 145 150 155 160
 Pro Val Ala Gly Pro Leu Gly Arg Ile Pro Glu Gly Gly Arg Asn Trp
 165 170 175
 Glu Gly Phe Ser Pro Asp Pro Val Leu Thr Gly Ile Ala Met Ala Glu
 180 185 190
 Thr Ile Lys Gly Met Gln Asp Thr Gly Val Ile Ala Cys Ala Lys His
 195 200 205
 Tyr Ile Gly Asn Glu Gln Glu His Phe Arg Gln Val Gly Glu Ala Ala
 210 215 220
 Gly His Gly Tyr Thr Ile Ser Asp Thr Ile Ser Ser Asn Ile Asp Asp
 225 230 235 240
 Arg Ala Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg
 245 250 255
 Ala Gly Val Gly Ser Phe Met Cys Ser Tyr Ser Gln Ile Asn Asn Ser
 260 265 270
 Tyr Gly Cys Gln Asn Ser Gln Thr Leu Asn Lys Leu Leu Lys Ser Glu
 275 280 285
 Leu Gly Phe Gln Gly Phe Val Met Ser Asp Trp Gly Ala His His Ser
 290 295 300

Gly Val Ser Ser Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp
 305 310 315 320
 Thr Glu Phe Asp Ser Gly Leu Ser Phe Trp Gly Ser Asn Leu Thr Ile
 325 330 335
 Ala Ile Leu Asn Gly Thr Val Pro Glu Trp Arg Leu Asp Asp Met Ala
 340 345 350
 Met Arg Ile Met Ala Ala Tyr Phe Lys Val Gly Leu Thr Ile Glu Asp
 355 360 365
 Gln Pro Asp Val Asn Phe Asn Ala Trp Thr His Asp Thr Tyr Gly Tyr
 370 375 380
 Lys Tyr Ala Tyr Ser Lys Glu Asp Tyr Glu Gln Val Asn Trp His Val
 385 390 395 400
 Asp Val Arg Ser Asp His Asn Lys Leu Ile Arg Glu Thr Ala Ala Lys
 405 410 415
 Gly Thr Val Leu Leu Lys Asn Asn Phe His Ala Leu Pro Leu Lys Gln
 420 425 430
 Pro Arg Phe Val Ala Val Val Gly Gln Asp Ala Gly Pro Asn Pro Lys
 435 440 445
 Gly Pro Asn Gly Cys Ala Asp Arg Gly Cys Asp Gln Gly Thr Leu Ala
 450 455 460
 Met Gly Trp Gly Ser Gly Ser Thr Glu Phe Pro Tyr Leu Val Thr Pro
 465 470 475 480
 Asp Thr Ala Ile Gln Ser Lys Val Leu Glu Tyr Gly Gly Arg Tyr Glu
 485 490 495
 Ser Ile Phe Asp Asn Tyr Asp Asp Asn Ala Ile Leu Ser Leu Val Ser
 500 505 510
 Gln Pro Asp Ala Thr Cys Ile Val Phe Ala Asn Ala Asp Ser Gly Glu
 515 520 525
 Gly Tyr Ile Thr Val Asp Asn Asn Trp Gly Asp Arg Asn Asn Leu Thr
 530 535 540
 Leu Trp Gln Asn Ala Asp Gln Val Ile Ser Thr Val Ser Ser Arg Cys
 545 550 555 560
 Asn Asn Thr Ile Val Val Leu His Ser Val Gly Pro Val Leu Leu Asn
 565 570 575
 Gly Ile Tyr Glu His Pro Asn Ile Thr Ala Ile Val Trp Ala Gly Met
 580 585 590
 Pro Gly Glu Glu Ser Gly Asn Ala Leu Val Asp Ile Leu Trp Gly Asn
 595 600 605
 Val Asn Pro Ala Gly Arg Thr Pro Phe Thr Trp Ala Lys Ser Arg Glu
 610 615 620
 Asp Tyr Gly Thr Asp Ile Met Tyr Glu Pro Asn Asn Gly Gln Arg Ala
 625 630 635 640
 Pro Gln Gln Asp Phe Thr Glu Ser Ile Tyr Leu Asp Tyr Arg His Phe
 645 650 655
 Asp Lys Ala Gly Ile Glu Pro Ile Tyr Glu Phe Gly Phe Gly Leu Ser
 660 665 670
 Tyr Thr Thr Phe Glu Tyr Ser Asp Leu Arg Val Val Lys Lys Tyr Val
 675 680 685
 Gln Pro Tyr Ser Pro Thr Thr Gly Thr Gly Ala Gln Ala Pro Ser Ile
 690 695 700
 Gly Gln Pro Pro Ser Gln Asn Leu Asp Thr Tyr Lys Phe Pro Ala Thr
 705 710 715 720
 Tyr Lys Tyr Ile Lys Thr Phe Ile Tyr Pro Tyr Leu Asn Ser Thr Val
 725 730 735
 Ser Leu Arg Ala Ala Ser Lys Asp Pro Glu Tyr Gly Arg Thr Asp Phe
 740 745 750
 Ile Pro Pro His Ala Arg Asp Gly Ser Pro Gln Pro Leu Asn Pro Ala
 755 760 765
 Gly Asp Pro Val Ala Ser Gly Gly Asn Asn Met Leu Tyr Asp Glu Leu
 770 775 780
 Tyr Glu Val Thr Ala Gln Ile Lys Asn Thr Gly Asp Val Ala Gly Asp
 785 790 795 800
 Glu Val Val Gln Leu Tyr Val Asp Leu Gly Gly Asp Asn Pro Pro Arg

Ala Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val Ser
 50 55 60
 Gly Val Gly Trp Asn Gly Gly Pro Cys Val Gly Asn Thr Ser Pro Ala
 65 70 75 80
 Ser Lys Ile Ser Tyr Pro Ser Leu Cys Leu Gln Asp Gly Pro Leu Gly
 85 90 95
 Val Arg Tyr Ser Thr Gly Ser Thr Ala Phe Thr Pro Gly Val Gln Ala
 100 105 110
 Ala Ser Thr Trp Asp Val Asn Leu Ile Arg Glu Arg Gly Gln Phe Ile
 115 120 125
 Gly Glu Glu Val Lys Ala Ser Gly Ile His Val Ile Leu Gly Pro Val
 130 135 140
 Ala Gly Pro Leu Gly Lys Thr Pro Gln Gly Gly Arg Asn Trp Glu Gly
 145 150 155 160
 Phe Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr Ile
 165 170 175
 Asn Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr Ile
 180 185 190
 Leu Asn Glu Gln Glu Leu Asn Arg Glu Thr Ile Ser Ser Asn Pro Asp
 195 200 205
 Asp Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala Val
 210 215 220
 Gln Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn Thr
 225 230 235 240
 Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys Asp
 245 250 255
 Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln His
 260 265 270
 Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro Gly
 275 280 285
 Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr Asn
 290 295 300
 Ala Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met Val
 305 310 315 320
 Thr Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala Gly
 325 330 335
 Tyr Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys Thr
 340 345 350
 Asn Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn Asp
 355 360 365
 Ala Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val Gly
 370 375 380
 Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn
 385 390 395 400
 Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly
 405 410 415
 Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr
 420 425 430
 Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn
 435 440 445
 Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val
 450 455 460
 Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn
 465 470 475 480
 Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu
 485 490 495
 Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His
 500 505 510
 Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val
 515 520 525
 Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala
 530 535 540
 Leu Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val

<212> DNA
 <213> *Aspergillus niger*
 <400> 39
 gtactaaaac c 11
 <210> 40
 <211> 11
 <212> DNA
 <213> *Aspergillus niger*
 <400> 40
 ccgttaaatt t 11
 <210> 41
 <211> 45
 <212> DNA
 <213> *Aspergillus niger*
 <400> 41
 ggatgctggt gactccggaa atttaacggt ttggtcttgc atccc 45
 <210> 42
 <211> 14
 <212> DNA
 <213> *Aspergillus niger*
 <400> 42
 atgcaattta aact 14
 <210> 43
 <211> 14
 <212> DNA
 <213> *Aspergillus niger*
 <400> 43
 cggcaattta acgg 14
 <210> 44
 <211> 44
 <212> DNA
 <213> *Aspergillus niger*
 <400> 44
 ggtattgtcc tgcagacggc aatttaacgg cttctgcgaa togc 44
 <210> 45
 <211> 29
 <212> DNA
 <213> *Humicola insolens*
 <400> 45
 aagcttaagc atgcgcttcc cccccctcc 29
 <210> 46
 <211> 32
 <212> DNA
 <213> *Humicola insolens*
 <400> 46
 ctgcagaatt ctacaggcac tgatggtacc ag 32
 <210> 47
 <211> 32
 <212> DNA
 <213> *Trichoderma reesei*
 <400> 47
 ctgcagaatt ctacaggcac tgatggtacc ag 32
 <210> 48
 <211> 36
 <212> DNA
 <213> *Trichoderma reesei*
 <400> 48
 accgcggact gcgcatcatg cgttcctccc ccctcc 36
 <210> 49
 <211> 29
 <212> DNA
 <213> *Trichoderma reesei*
 <400> 49

aaacgtcgac cgaatgtagg attgttatc	29
<210> 50	
<211> 17	
<212> DNA	
<213> Trichoderma reesei	
<400> 50	
gatgcgcagt ccgcggt	17
<210> 51	
<211> 29	
<212> DNA	
<213> Trichoderma reesei	
<400> 51	
aaacgtcgac cgaatgtagg attgttatc	29
<210> 52	
<211> 36	
<212> DNA	
<213> Trichoderma reesei	
<400> 52	
ggagggggga ggaacgcatg atgcgagtc cgcggt	36
<210> 53	
<211> 29	
<212> DNA	
<213> Trichoderma reesei	
<400> 53	
aaacgtcgac cgaatgtagg attgttatc	29
<210> 54	
<211> 32	
<212> DNA	
<213> Trichoderma reesei	
<400> 54	
ctgcagaatt ctacaggcac tgatggtacc ag	32
<210> 55	
<211> 46	
<212> DNA	
<213> Aspergillus oryzae	
<400> 55	
atagtcaacc gcggactgcg catcatgaag cttggttgga tcgagg	46
<210> 56	
<211> 26	
<212> DNA	
<213> Aspergillus oryzae	
<400> 56	
actagtttac tgggccttag gcagcg	26
<210> 57	
<211> 26	
<212> DNA	
<213> Trichoderma reesei	
<400> 57	
gtcgactcga agcccgaatg taggat	26
<210> 58	
<211> 45	
<212> DNA	
<213> Trichoderma reesei	
<400> 58	
cctcgatcca accaagcttc atgatgcgca gtccgcggtt gacta	45
<210> 59	
<211> 57	
<212> DNA	
<213> Aspergillus oryzae	
<400> 59	
atgaagcttg gttggatcga ggtggccgca ttggcggctg cctcagtagt cagtgcc	57
<210> 60	
<211> 19	

<400> 69
 gccgtccaga tccccatgcg ttcttcccc 30
 <210> 70
 <211> 20
 <212> DNA
 <213> *Aspergillus oryzae*
 <400> 70
 ccaagcttgt tcagagtttc 20
 <210> 71
 <211> 20
 <212> DNA
 <213> *Aspergillus oryzae*
 <400> 71
 ggactgcgca gcatgcggttc 20
 <210> 72
 <211> 30
 <212> DNA
 <213> *Aspergillus oryzae*
 <400> 72
 agttaattaa ttactgggccc ttaggcagcgc 30
 <210> 73
 <211> 3294
 <212> DNA
 <213> *Aspergillus oryzae*
 <400> 73
 atgcgttctt cccccctctt ccgctccgccc gttgtggccg ccctgcccgt gttggccctt 60
 gccgctgatg gcaggtccac ccgctactgg gactgctgca agccttcgtg cggctgggcc 120
 aagaaggctc ccgtgaacca gcctgtcttt tcctgcaacg ccaacttcca gcgtatcacg 180
 gacttcgacg ccaagtccgg ctgcgagcgg ggccggtgctg cctactcgtg cgccgaccag 240
 accccatggg ctgtgaacga cgacttcgcg ctcggttttg ctgccacctc tattgcccggc 300
 agcaatgagg cgggctgggt ctgcccctgc tacgagctca ccttcacatc cggctctggt 360
 gctggcaaga agatggctgt ccagtcaccc agcactggcg gtgatcttgg cagcaaccac 420
 ttcgatctca acatccccgg cggcggcgctc ggcactcttcg acggatgcac tccccagttc 480
 ggtggtctgc ccggccagcg ctacggcggc atctcgtccc gcaacgagtg cgatcggttc 540
 cccgacgccc tcaagcccgg ctgctactgg cgcttcgact ggttcaagaa cgccgacaat 600
 ccgagcttca gcttccgctca ggtccagtg ccagccgagc tcgtcgtctg caccggatgc 660
 cgccgcaacg acgacggcaa cttccctgcc gtcacagatcc ccattgcgttc ctccccctc 720
 ctccgctccg ccggttgggc cgccttgcgg gtggtggccc ttgccaagga tgatctcgcg 780
 tactccccct ctttctaccc ttccccatgg gcagatggtc aggggtgaatg ggccggaagta 840
 taaaaacgcg ctgtagacat agtttcccag atgacggtga cagagaaagt caacttaacg 900
 actggaacag gatggcaact agagaggtgt gttggacaaa ctggcagtggt tcccagactc 960
 aacatcccca gcttgtggtt gcaggatagt cctcttggta ttcggttctc ggactacaat 1020
 tcagctttcc ctgcccgggt taatgtcgtt gccacctggg acaagacgct cgcctacctt 1080
 cgtggctcagg caatgggtga ggagttcagt gataagggtg ttgacgttca gctgggtcct 1140
 gctgctggcc ctctcgggtg tcatccggat ggcggtgaga actgggaagg tttctcacca 1200
 gatccagccc tcaccggtgt acttttggcg gagacgatta agggattca agatgctggt 1260
 gtcattgcga cagctaagca ttatatcatg aacgaacaag agcatttccg ccaacaacc 1320
 gaggetgccc gttacggatt caacgtaagc gacagtttga gttccaactg tgatgacaag 1380
 actatgcatg aattgtacct ctggcccctc gcggatgcag tacgcgctgg agtcggtgct 1440
 gtcattgctt cttacaacca aatcaacaac agctacggtt gcgagaatag cgaaactctg 1500
 aacaagcttt tgaaggcggg gcttgggttc caaggcttcg tcatgagtga ttggaccgct 1560
 catcacagcg gcgtaggcgc tgcttttagca ggtctggata tgctgatgcc cgggtgatgtt 1620
 accttcgata gtggtacgtc tttctggggt gcaaacttga cggctcgggt ccttaacggt 1680
 acaatcccc aatggcgtgt tgatgacatg gctgtccgta tcatggccgc ttattacaag 1740
 gttggccgag acaccaata caccctccc aacttcagct cgtggaccag ggacgaatat 1800
 ggtttcgcgc ataaccatgt ttcggaaggt gcttacgaga gggtaacga attcgtggac 1860
 gtgcaacgcg atcatgccga cctaatecgt cgcacgcccg cgcagagcac tgttctgctg 1920
 aagaacaagg gtgccttggc cttgagcccg aaggaaaagc tggctgcctt tctgggagag 1980
 gatgcccgtt ccaactcgtg gggcgctaac ggctgtgatg accgtgggtg cgataacggt 2040
 acccttgcca tggcctgggg tagcggtaact gcgaatttcc catacctcgt gacaccagag 2100
 caggcgattc agaacgaagt tcttcagggc cgtggtaatg tcttcgcccgt gaccgacagt 2160
 tgggcgctcg acaagatcgc tgcggctgcc cgccaggcca gcgtatctct cgtgttctgc 2220
 aactccgact caggagaagg ctatcttagt gtggatggaa atgagggcga tcgtaacaac 2280

```

atcactctgt ggaagaacgg cgacaatgtg gtcaagaccg cagcgaataa ctgtaacaac 2340
accgtttgtca tcatccactc cgtcggacca gttttgatcg atgaatggta tgaccacccc 2400
aatgtcactg gtattctctg ggctggctg ccaggccagg agtctggtaa ctccattgcc 2460
gatgtgctgt acggctgtgt caaccctggc gccaaagtct ctttacttg gggcaagacc 2520
cgggagtcgt atggttctcc cttgggtcaag gatgccaaca atggcaacgg agcgccccag 2580
tctgatttca cccagggtgt tttcatcgat taccgccatt tcgataagtt caatgagacc 2640
cctatctacg agtttggcta cggcttgagc tacaccacct tcgagctctc cgacctccat 2700
gttcagcccc tgaacgcgtc ccgatacact cccaccagtg gcatgactga agctgcaaag 2760
aactttgggtg aaattggcga tgcgtcggag tacgtgtatc cggaggggct ggaaggatc 2820
catgagttta tctatccctg gatcaactct accgacctga aggcacgctc tgacgattct 2880
aactacggct ggggaagactc caagtatatt cccgaaggcg ccacggatgg gtctgcccag 2940
ccccgtttgc ccgctagtgg tggtgccgga ggaaaccccg gtctgtacga ggatcttttc 3000
cgcgtctctg tgaaggtcaa gaacacgggc aatgtcgcg gtgatgaagt tcctcagctg 3060
tacgtttccc taggcggccc gaatgagccc aagggtgtac tgcgcaagtt tgagcgtatt 3120
cacttggccc cttcgcagga ggccgtgtgg acaacgaccc ttaccctgctg tgaccttgca 3180
aactgggacg tttcggctca ggactggacc gtcactcctt accccaagac gatctacgtt 3240
ggaaactcct cacggaaact gccgctccag gcctcgtctg ctaaggccca gtaa 3294

```

<210> 74

<211> 1097

<212> PRT

<213> *Aspergillus oryzae*

<400> 74

```

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro
1          5          10          15
Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
20          25          30
Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
35          40          45
Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
50          55          60
Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
65          70          75          80
Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
85          90          95
Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
100         105         110
Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
115         120         125
Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
130         135         140
Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
145         150         155         160
Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
165         170         175
Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
180         185         190
Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
195         200         205
Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
210         215         220
Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Met Arg Ser Ser Pro Leu
225         230         235         240
Leu Arg Ser Ala Val Val Ala Ala Leu Pro Val Leu Ala Leu Ala Lys
245         250         255
Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser Pro Trp Ala Asp
260         265         270
Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala Val Asp Ile Val
275         280         285
Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr Thr Gly Thr Gly
290         295         300
Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser Val Pro Arg Leu
305         310         315         320
Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu Gly Ile Arg Phe

```

				325					330				335			
Ser	Asp	Tyr	Asn	Ser	Ala	Phe	Pro	Ala	Gly	Val	Asn	Val	Ala	Ala	Thr	
			340					345					350			
Trp	Asp	Lys	Thr	Leu	Ala	Tyr	Leu	Arg	Gly	Gln	Ala	Met	Gly	Glu	Glu	
		355					360					365				
Phe	Ser	Asp	Lys	Gly	Ile	Asp	Val	Gln	Leu	Gly	Pro	Ala	Ala	Gly	Pro	
		370				375					380					
Leu	Gly	Ala	His	Pro	Asp	Gly	Gly	Arg	Asn	Trp	Glu	Gly	Phe	Ser	Pro	
385					390					395					400	
Asp	Pro	Ala	Leu	Thr	Gly	Val	Leu	Phe	Ala	Glu	Thr	Ile	Lys	Gly	Ile	
				405					410					415		
Gln	Asp	Ala	Gly	Val	Ile	Ala	Thr	Ala	Lys	His	Tyr	Ile	Met	Asn	Glu	
			420					425					430			
Gln	Glu	His	Phe	Arg	Gln	Gln	Pro	Glu	Ala	Ala	Gly	Tyr	Gly	Phe	Asn	
		435					440					445				
Val	Ser	Asp	Ser	Leu	Ser	Ser	Asn	Val	Asp	Asp	Lys	Thr	Met	His	Glu	
		450				455					460					
Leu	Tyr	Leu	Trp	Pro	Phe	Ala	Asp	Ala	Val	Arg	Ala	Gly	Val	Gly	Ala	
465					470					475					480	
Val	Met	Cys	Ser	Tyr	Asn	Gln	Ile	Asn	Asn	Ser	Tyr	Gly	Cys	Glu	Asn	
				485				490						495		
Ser	Glu	Thr	Leu	Asn	Lys	Leu	Leu	Lys	Ala	Glu	Leu	Gly	Phe	Gln	Gly	
			500					505					510			
Phe	Val	Met	Ser	Asp	Trp	Thr	Ala	His	His	Ser	Gly	Val	Gly	Ala	Ala	
		515				520						525				
Leu	Ala	Gly	Leu	Asp	Met	Ser	Met	Pro	Gly	Asp	Val	Thr	Phe	Asp	Ser	
		530				535					540					
Gly	Thr	Ser	Phe	Trp	Gly	Ala	Asn	Leu	Thr	Val	Gly	Val	Leu	Asn	Gly	
545					550					555					560	
Thr	Ile	Pro	Gln	Trp	Arg	Val	Asp	Asp	Met	Ala	Val	Arg	Ile	Met	Ala	
				565					570					575		
Ala	Tyr	Tyr	Lys	Val	Gly	Arg	Asp	Thr	Lys	Tyr	Thr	Pro	Pro	Asn	Phe	
			580					585					590			
Ser	Ser	Trp	Thr	Arg	Asp	Glu	Tyr	Gly	Phe	Ala	His	Asn	His	Val	Ser	
		595				600						605				
Glu	Gly	Ala	Tyr	Glu	Arg	Val	Asn	Glu	Phe	Val	Asp	Val	Gln	Arg	Asp	
		610				615					620					
His	Ala	Asp	Leu	Ile	Arg	Arg	Ile	Gly	Ala	Gln	Ser	Thr	Val	Leu	Leu	
625					630					635					640	
Lys	Asn	Lys	Gly	Ala	Leu	Pro	Leu	Ser	Arg	Lys	Glu	Lys	Leu	Val	Ala	
				645					650					655		
Leu	Leu	Gly	Glu	Asp	Ala	Gly	Ser	Asn	Ser	Trp	Gly	Ala	Asn	Gly	Cys	
			660					665					670			
Asp	Asp	Arg	Gly	Cys	Asp	Asn	Gly	Thr	Leu	Ala	Met	Ala	Trp	Gly	Ser	
		675					680					685				
Gly	Thr	Ala	Asn	Phe	Pro	Tyr	Leu	Val	Thr	Pro	Glu	Gln	Ala	Ile	Gln	
		690				695					700					
Asn	Glu	Val	Leu	Gln	Gly	Arg	Gly	Asn	Val	Phe	Ala	Val	Thr	Asp	Ser	
705					710					715					720	
Trp	Ala	Leu	Asp	Lys	Ile	Ala	Ala	Ala	Ala	Arg	Gln	Ala	Ser	Val	Ser	
				725					730					735		
Leu	Val	Phe	Val	Asn	Ser	Asp	Ser	Gly	Glu	Gly	Tyr	Leu	Ser	Val	Asp	
			740					745					750			
Gly	Asn	Glu	Gly	Asp	Arg	Asn	Asn	Ile	Thr	Leu	Trp	Lys	Asn	Gly	Asp	
		755					760					765				
Asn	Val	Val	Lys	Thr	Ala	Ala	Asn	Asn	Cys	Asn	Asn	Thr	Val	Val	Ile	
		770				775					780					
Ile	His	Ser	Val	Gly	Pro	Val	Leu	Ile	Asp	Glu	Trp	Tyr	Asp	His	Pro	
785					790					795					800	
Asn	Val	Thr	Gly	Ile	Leu	Trp	Ala	Gly	Leu	Pro	Gly	Gln	Glu	Ser	Gly	
				805				810						815		
Asn	Ser	Ile	Ala	Asp	Val	Leu	Tyr	Gly	Arg	Val	Asn	Pro	Gly	Ala	Lys	
			820					825					830			

Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr Gly Ser Pro Leu
835 840 845
Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln Ser Asp Phe Thr
850 855 860
Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys Phe Asn Glu Thr
865 870 875 880
Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu Leu
885 890 895
Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg Tyr Thr Pro Thr
900 905 910
Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu Ile Gly Asp Ala
915 920 925
Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile His Glu Phe Ile
930 935 940
Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser Ser Asp Asp Ser
945 950 955 960
Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu Gly Ala Thr Asp
965 970 975
Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly Ala Gly Gly Asn
980 985 990
Pro Gly Leu Tyr Glu Asp Leu Phe Arg Val Ser Val Lys Val Lys Asn
995 1000 1005
Thr Gly Asn Val Ala Gly Asp Glu Val Pro Gln Leu Tyr Val Ser
1010 1015 1020
Leu Gly Gly Pro Asn Glu Pro Lys Val Val Leu Arg Lys Phe Glu
1025 1030 1035
Arg Ile His Leu Ala Pro Ser Gln Glu Ala Val Trp Thr Thr Thr
1040 1045 1050
Leu Thr Arg Arg Asp Leu Ala Asn Trp Asp Val Ser Ala Gln Asp
1055 1060 1065
Trp Thr Val Thr Pro Tyr Pro Lys Thr Ile Tyr Val Gly Asn Ser
1070 1075 1080
Ser Arg Lys Leu Pro Leu Gln Ala Ser Leu Pro Lys Ala Gln
1085 1090 1095

<210> 75

<211> 3294

<212> DNA

<213> *Aspergillus oryzae*

<400> 75

atgcggttct cccccctcct ccgctccgcc gttgtggccg cctgcccgtt gttggccctt 60
gccgctgatg gcaggtccac ccgctactgg gactgctgca agccttcgtg cggctggggc 120
aagaaggctc ccgtgaacca gcctgtcttt tcctgcaacg ccaacttcca gcgtatcacg 180
gacttcgacg ccaagtcggg ctgcgagccg ggccggtgtg cctactcgtg cgccgaccag 240
accccatggg ctgtgaacga cgacttcgag ctccggtttg ctgccacctc tattgccggc 300
agcaatgagc cgggctggtg ctgcccctgc tacgagctca ccttcacatc cggctctggt 360
gctggcaaga agatggtcgt ccagtcacc agcactggcg gtgatcttgg cagcaaccac 420
ttcgatctca acatccccgg cggcggcgct gccatcttcg acggatgcac tccccagttc 480
ggtggtctgc ccggccagcg ctacggcggc atctcgtccc gcaacgagtg cgatcggttc 540
cccagcggcc tcaagccggg ctgctactgg cgcttcgact ggttcaagaa cgccgacaat 600
ccgagcttca gcttccgtca ggtccagtg ccagccgagc tcgtcgtctg caccggatgc 660
cgccgcaacg acgacggcaa ctccctgccc gtccagatcc ccatgcgttc ctccccctc 720
ctccgctccg ccggtgtggc cgcctgccc gtgttggccc ttgccaagga tgatctcgcg 780
tactcccctc ctttctaccc ttccccatgg gcagatggtg aggggtgaatg ggcggaagta 840
tacaaacgcg ctgtagacat agtttcccag atgacgttga cagagaaagt caacttaacg 900
actggaacag gatggcaact agagaggtgt gttggacaaa ctggcagtg tcccagactc 960
aacatcccga gcttgtgttt gcaggatagt cctcttggtt ttcgtttctc ggactacaat 1020
tcagctttcc ctgcccgtgt taatgtcgt gccacctggg acaagacgt cgcctacctt 1080
cgtggctcagg caatgggtga ggagttcagt gataagggtt ttgacgttca gctgggtcct 1140
gctgctggcc ctctcgggtg tcatccggat ggccgtagaa actgggaaag tttctacca 1200
gatccagccc tcaccgggtg actttttgcg gagacgatta agggatttca agatgctggt 1260
gtcattgcga cagctaagca ttatatcatg aacgaacaag agcatttccg ccaacaaccc 1320
gaggctgagg gttacggatt caacgtaagc gacagtttga gttccaacgt tgatgacaag 1380
actatgcatg aattgtacct ctggcccttc gcggatgcag tacgcgctgg agtcggtgct 1440

gttatgtgct	cttacaacca	aatcaacaac	agctacggtt	gcgagaatag	cgaaactctg	1500									
aacaagcttt	tgaaggcgga	gcttggtttc	caaggcttcg	tcatgagtga	ttggaccgct	1560									
caacacagcg	gcgtaggcgc	tgctttagca	ggtctggata	tgctgatgcc	cggtgatggt	1620									
accttcgata	gtggtacgtc	tttctgggg	gcaaacttga	cggtcgggtg	ccttaacggt	1680									
acaatcccc	aatggcgtgt	tgatgacatg	gctgtccgta	tcatggccgc	ttattacaag	1740									
gttggccgcg	acaccaaata	caccctccc	aacttcagct	cgtggaccag	ggacgaatat	1800									
ggtttcgcgc	ataaccatgt	ttcgggaagt	gcttacgaga	gggtcaacga	attcgtggac	1860									
gtgcaacgcg	atcatgccga	cctaaccctg	cgcatcggcg	cgagagcac	tgttctgctg	1920									
aagaacaagg	gtgccttgcc	cttgagccgc	aaggaaaagc	tggtcgcct	tctgggagag	1980									
gatgcggtt	ccaactcgtg	gggcgctaac	ggctgtgatg	accgtggtt	cgataacggt	2040									
acccttgcca	tggcctgggg	tagcggta	gcgaatttcc	catacctcgt	gacaccagag	2100									
caggcgattc	agaacgaagt	tcttcagggc	cgtggtaatg	tcttcgccgt	gaccgacagt	2160									
tgggcgctcg	acaagatcgc	tgcggctgcc	cgccaggcca	gcgtatctct	cgtgttcgct	2220									
aactccgact	caggagaagg	ctatcttagt	gtggatggaa	atgaggcgca	tcgtaacaac	2280									
atcactctgt	ggaagaacgg	cgacaatgtg	gtcaagaccg	cagcgaataa	ctgtaacaac	2340									
accgttgtca	tcattccactc	cgctcgacca	gttttgatcg	atgaatggta	tgaccacccc	2400									
aatgtcactg	gtattctctg	ggctggctcg	ccaggccagg	agtctggtaa	ctccattgcc	2460									
gatgtgctgt	acggctcgtg	caaccctggc	gccaagtctc	ctttcacttg	gggcaagacc	2520									
cgggagtogt	atggttctcc	cttgggtcaag	gatgccaaaca	atggcaacgg	agcgcaccag	2580									
tctgatttca	cccagggtgt	tttcatcgat	taccgccatt	tcgataagtt	caatgagacc	2640									
cctatctacg	agtttggcta	cggttgagc	tacaccacct	tcgagctctc	cgacctccat	2700									
gttcagcccc	tgaacgcgctc	ccgatacact	cccaccagtg	gcatgactga	agctgcaaag	2760									
aactttggtg	aaattggcga	tgcgctcggag	tacgtgtatc	cggaggggct	ggaaaggatc	2820									
catgagttta	tctatccctg	gatcaactct	accgacctga	aggcatcgtc	tgacgattct	2880									
aactacggct	gggaagactc	caagtatatt	cccgaaggcg	ccacggatgg	gtctgcccag	2940									
ccccgtttgc	ccgctagtgg	tggtgccgga	ggaaaccccg	gtctgtacga	ggatcttttc	3000									
cgcgtctctg	tgaaggtcaa	gaacacgggc	aatgtcggcg	gtgatgaagt	tcctcacttg	3060									
tacgtttccc	taggcggccc	gaatgagccc	aagggtgtac	tgcgcaagtt	tgagcgtatt	3120									
cacttggccc	cttcgcagga	ggccgtgtgg	acaacgaccc	ttaccctcgt	tgaccttgca	3180									
aactgggacg	tttcggctca	ggactggacc	gtcactcctt	accccaagac	gatctacggt	3240									
ggaaactcct	cacggaaact	gcccgtccag	gcctcgtctg	ctaaggccca	gtaa	3294									
<210>	76														
<211>	1097														
<212>	PRT														
<213>	Aspergillus oryzae														
<400>	76														
Met	Arg	Ser	Ser	Pro	Leu	Leu	Arg	Ser	Ala	Val	Val	Ala	Ala	Leu	Pro
1			5					10						15	
Val	Leu	Ala	Leu	Ala	Ala	Asp	Gly	Arg	Ser	Thr	Arg	Tyr	Trp	Asp	Cys
			20					25					30		
Cys	Lys	Pro	Ser	Cys	Gly	Trp	Ala	Lys	Lys	Ala	Pro	Val	Asn	Gln	Pro
		35					40					45			
Val	Phe	Ser	Cys	Asn	Ala	Asn	Phe	Gln	Arg	Ile	Thr	Asp	Phe	Asp	Ala
	50					55					60				
Lys	Ser	Gly	Cys	Glu	Pro	Gly	Gly	Val	Ala	Tyr	Ser	Cys	Ala	Asp	Gln
65					70					75					80
Thr	Pro	Trp	Ala	Val	Asn	Asp	Asp	Phe	Ala	Leu	Gly	Phe	Ala	Ala	Thr
				85					90						95
Ser	Ile	Ala	Gly	Ser	Asn	Glu	Ala	Gly	Trp	Cys	Cys	Ala	Cys	Tyr	Glu
			100					105					110		
Leu	Thr	Phe	Thr	Ser	Gly	Pro	Val	Ala	Gly	Lys	Lys	Met	Val	Val	Gln
		115					120					125			
Ser	Thr	Ser	Thr	Gly	Gly	Asp	Leu	Gly	Ser	Asn	His	Phe	Asp	Leu	Asn
		130				135					140				
Ile	Pro	Gly	Gly	Gly	Val	Gly	Ile	Phe	Asp	Gly	Cys	Thr	Pro	Gln	Phe
145					150					155					160
Gly	Gly	Leu	Pro	Gly	Gln	Arg	Tyr	Gly	Gly	Ile	Ser	Ser	Arg	Asn	Glu
				165					170					175	
Cys	Asp	Arg	Phe	Pro	Asp	Ala	Leu	Lys	Pro	Gly	Cys	Tyr	Trp	Arg	Phe
			180					185					190		
Asp	Trp	Phe	Lys	Asn	Ala	Asp	Asn	Pro	Ser	Phe	Ser	Phe	Arg	Gln	Val
		195					200					205			
Gln	Cys	Pro	Ala	Glu	Leu	Val	Ala	Arg	Thr	Gly	Cys	Arg	Arg	Asn	Asp

Trp Ala Leu Asp Lys Ile Ala Ala Ala Ala Arg Gln Ala Ser Val Ser
 725 730 735
 Leu Val Phe Val Asn Ser Asp Ser Gly Glu Gly Tyr Leu Ser Val Asp
 740 745 750
 Gly Asn Glu Gly Asp Arg Asn Asn Ile Thr Leu Trp Lys Asn Gly Asp
 755 760 765
 Asn Val Val Lys Thr Ala Ala Asn Asn Cys Asn Asn Thr Val Val Ile
 770 775 780
 Ile His Ser Val Gly Pro Val Leu Ile Asp Glu Trp Tyr Asp His Pro
 785 790 795 800
 Asn Val Thr Gly Ile Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly
 805 810 815
 Asn Ser Ile Ala Asp Val Leu Tyr Gly Arg Val Asn Pro Gly Ala Lys
 820 825 830
 Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr Gly Ser Pro Leu
 835 840 845
 Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln Ser Asp Phe Thr
 850 855 860
 Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys Phe Asn Glu Thr
 865 870 875 880
 Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu Leu
 885 890 895
 Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg Tyr Thr Pro Thr
 900 905 910
 Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu Ile Gly Asp Ala
 915 920 925
 Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile His Glu Phe Ile
 930 935 940
 Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser Ser Asp Asp Ser
 945 950 955 960
 Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu Gly Ala Thr Asp
 965 970 975
 Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly Ala Gly Gly Asn
 980 985 990
 Pro Gly Leu Tyr Glu Asp Leu Phe Arg Val Ser Val Lys Val Lys Asn
 995 1000 1005
 Thr Gly Asn Val Ala Gly Asp Glu Val Pro Gln Leu Tyr Val Ser
 1010 1015 1020
 Leu Gly Gly Pro Asn Glu Pro Lys Val Val Leu Arg Lys Phe Glu
 1025 1030 1035
 Arg Ile His Leu Ala Pro Ser Gln Glu Ala Val Trp Thr Thr Thr
 1040 1045 1050
 Leu Thr Arg Arg Asp Leu Ala Asn Trp Asp Val Ser Ala Gln Asp
 1055 1060 1065
 Trp Thr Val Thr Pro Tyr Pro Lys Thr Ile Tyr Val Gly Asn Ser
 1070 1075 1080
 Ser Arg Lys Leu Pro Leu Gln Ala Ser Leu Pro Lys Ala Gln
 1085 1090 1095

REIVINDICAÇÕES

1. Método para produzir um polipeptídeo secretado tendo atividade biológica, caracterizado pelo fato de compreender:

5 (a) transformar uma célula hospedeira fúngica com uma construção de proteína de fusão codificando uma proteína de fusão, em que a construção de proteína de fusão compreende:

(i) um primeiro polinucleotídeo compreendendo uma sequência nucleotídica codificando um peptídeo de sinal;

10 (ii) um segundo polinucleotídeo codificando uma sequência nucleotídica codificando pelo menos um domínio catalítico de uma endoglucanase ou uma porção da mesma; e

(iii) um terceiro polinucleotídeo compreendendo uma sequência nucleotídica codificando pelo menos um domínio catalítico de um polipeptídeo tendo atividade biológica ou uma porção da mesma;

15 em que o peptídeo de sinal e pelo menos o domínio catalítico da endoglucanase ou da porção da mesma aumenta a secreção do polipeptídeo tendo atividade biológica ou a porção da mesma comparada à ausência de pelo menos o domínio catalítico da endoglucanase ou a porção da mesma;

20 (b) cultivar a célula hospedeira fúngica transformada sob condições apropriadas para a produção da proteína de fusão, e

(c) recuperar a proteína de fusão, um componente da mesma, ou uma combinação da mesma, do meio de cultivo, em que a proteína de fusão ou o componente da mesma tem atividade biológica.

25 2. Método de acordo com a reivindicação 1, caracterizado pelo fato de que a extremidade 3' do primeiro polinucleotídeo é ligada operativamente à extremidade 5' do segundo polinucleotídeo e a extremidade 3' do segundo polinucleotídeo é ligada operativamente à extremidade 5' do terceiro polinucleotídeo ou a extremidade 3' do primeiro polinucleotídeo é ligada operativamente à extremidade 5' do terceiro polinucleotídeo e a

extremidade 3' do terceiro polinucleotídeo é ligada operativamente à extremidade 5' do segundo polinucleotídeo para codificar uma proteína de fusão.

3. Método de acordo com a reivindicação 1 ou 2, caracterizado pelo fato de que o primeiro polinucleotídeo codificando o peptídeo de sinal é obtido de um gene de endoglucanase CEL45 de *Humicola insolens* (*cel45*).

4. Método de acordo com qualquer uma das reivindicações 1-3, caracterizado pelo fato de que o segundo polinucleotídeo compreendendo a sequência nucleotídica codificando pelo menos o domínio catalítico de endoglucanase ou a porção do mesmo é obtido a partir de um gene de endoglucanase CEL45 de *Humicola insolens* (*cel45*).

5. Método de acordo com qualquer uma das reivindicações 1-3, caracterizado pelo fato de que o segundo polinucleotídeo compreendendo a sequência nucleotídica codificando pelo menos o domínio catalítico de endoglucanase ou a porção do mesmo é obtido de um gene codificando uma endoglucanase compreendendo uma sequência de aminoácido tendo pelo menos 70%, preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 85%, o mais preferivelmente pelo menos 90%, e ainda mais preferivelmente pelo menos 95% de identidade para SEQ ID NO: 2.

6. Método de acordo com qualquer uma das reivindicações 1-5, caracterizado pelo fato de que a construção de proteína de fusão ainda compreende um polinucleotídeo codificando um módulo de ligação de carboidrato.

7. Método de acordo com qualquer uma das reivindicações 1-6, caracterizado pelo fato de que o segundo polinucleotídeo codifica um domínio catalítico ou uma porção do mesmo, um polipeptídeo maduro, ou um polipeptídeo de comprimento completo da endoglucanase.

8. Método de acordo com qualquer uma das reivindicações 1-

7, caracterizado pelo fato de que o terceiro polinucleotídeo codifica um domínio catalítico ou uma porção do mesmo, um polipeptídeo maduro, ou um polipeptídeo de comprimento completo do polipeptídeo tendo atividade biológica.

5 9. Método de acordo com qualquer uma das reivindicações 1-
8, caracterizado pelo fato de que a construção de proteína de fusão ainda
compreende um polinucleotídeo codificando um ligando localizado entre a
sequência nucleotídica codificando pelo menos o domínio catalítico de
10 endoglucanase ou a porção do mesmo e a sequência nucleotídica codificando
pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica
ou a porção do mesmo e alternativamente também um polinucleotídeo
codificando um sítio de clivagem localizado entre a sequência nucleotídica
codificando pelo menos o domínio catalítico de endoglucanase ou a porção do
15 mesmo e a sequência nucleotídica codificando pelo menos o domínio
catalítico de um polipeptídeo tendo atividade biológica ou a porção do
mesmo.

 10. Método de acordo com qualquer uma das reivindicações 1-
9, caracterizado pelo fato de que a construção de proteína de fusão ainda
compreende um quarto polinucleotídeo codificando um segundo peptídeo de
20 sinal.

 11. Método de acordo com a reivindicação 10, caracterizado
pelo fato de que a extremidade 3' do quarto polinucleotídeo é ligado
operativamente para a extremidade 5' do segundo polinucleotídeo ou o
terceiro polinucleotídeo.

25 12. Método de acordo com qualquer uma das reivindicações 1-
11, caracterizado pelo fato de que a construção de proteína de fusão ainda
compreende um promotor e/ou terminador.

 13. Método de acordo com a reivindicação 12, caracterizado
pelo fato que o promotor aciona a expressão do primeiro, segundo e terceiro

polinucleotídeos, e alternativamente também do quarto polinucleotídeo.

14. Método de acordo com qualquer uma das reivindicações 1-13, caracterizado pelo fato de que o polipeptídeo tendo atividade biológica é um anticorpo, antígeno, peptídeo antimicrobiano, enzima, fator de crescimento, hormônio, imunodilatador, neurotransmissor, receptor, proteína de repórter, ou proteína estrutural.

15. Método de acordo com qualquer uma das reivindicações 1-14, caracterizado pelo fato de que o polipeptídeo tendo atividade biológica é uma oxidoreductase, transferase, hidrolase, liase, isomerase ou ligase.

16. Método de acordo com qualquer uma das reivindicações 1-15, caracterizado pelo fato de que o polipeptídeo tendo atividade biológica é uma enzima celulolítica ou uma enzima hemicelulolítica.

17. Método de acordo com a reivindicação 16, caracterizado pelo fato de que a enzima celulolítica é uma beta-glucosidase.

18. Método de acordo com a reivindicação 17, caracterizado pelo fato de que a beta-glucosidase compreende uma sequência de aminoácido tendo pelo menos 70%, preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 85%, o mais preferivelmente pelo menos 90%, e ainda o mais preferivelmente pelo menos 95% de identidade para SEQ ID NO: 24, ou uma beta-glucosidase variante compreendendo uma sequência de aminoácido tendo pelo menos 70%, preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 85%, o mais preferivelmente pelo menos 90%, e ainda o mais preferivelmente pelo menos 95% de identidade para SEQ ID NO: 35.

19. Método de acordo com qualquer uma das reivindicações 1-18, caracterizado pelo fato de que a proteína de fusão, um componente da mesma, ou uma combinação da mesma, tem atividade celulolítica ou atividade hemicelulolítica.

20. Proteína de fusão isolada, caracterizada pelo fato de compreender:

(a) uma primeira sequência de aminoácido compreendendo um peptídeo de sinal;

5 (b) uma segunda sequência de aminoácido compreendendo pelo menos um domínio catalítico de uma endoglucanase ou uma porção da mesma; e

(c) uma terceira sequência de aminoácido compreendendo pelo menos um domínio catalítico de um polipeptídeo tendo atividade biológica ou
10 uma porção da mesma.

21. Proteína de fusão de acordo com a reivindicação 20, caracterizada pelo fato de que a extremidade C-terminal da primeira sequência de aminoácido é ligada em armação à extremidade N-terminal da segunda sequência de aminoácido e a extremidade C-terminal da segunda
15 sequência de aminoácido é ligada em armação à extremidade N-terminal da terceira sequência de aminoácido ou a extremidade C-terminal da primeira sequência de aminoácido é ligada em armação à extremidade N-terminal da terceira sequência de aminoácido e a extremidade C-terminal da terceira
20 sequência de aminoácido é ligada em armação à extremidade N-terminal da segunda sequência de aminoácido.

22. Proteína de fusão de acordo com a reivindicação 20 ou 21, caracterizada pelo fato de que a primeira sequência de aminoácido compreendendo o peptídeo de sinal é obtida de endoglucanase CEL45 de *Humicola insolens*.

25 23. Proteína de fusão de acordo com qualquer uma das reivindicações 20-22, caracterizada pelo fato de que a segunda sequência de aminoácido compreendendo pelo menos o domínio catalítico de endoglucanase é obtida de uma endoglucanase CEL45 de *Humicola insolens*.

24. Proteína de fusão de acordo com qualquer uma das

reivindicações 20-22, caracterizada pelo fato de que a segunda sequência de aminoácido compreendendo pelo menos o domínio catalítico de endoglucanase é obtida de uma endoglucanase compreendendo uma sequência de aminoácido tendo pelo menos 70%, preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 85%, o mais preferivelmente pelo menos 90%, e ainda mais preferivelmente pelo menos 95% de identidade para SEQ ID NO: 2.

25. Proteína de fusão de acordo com qualquer uma das reivindicações 20-24, caracterizada pelo fato de ainda compreender um módulo de ligação de carboidrato.

26. Proteína de fusão de acordo com qualquer uma das reivindicações 20-25, caracterizada pelo fato de que a segunda sequência de aminoácido compreende um domínio catalítico ou uma porção do mesmo, um polipeptídeo maduro, ou um polipeptídeo de comprimento completo da endoglucanase.

27. Proteína de fusão de acordo com qualquer uma das reivindicações 20-26, caracterizada pelo fato de que a terceira sequência de aminoácido compreende um domínio catalítico ou uma porção do mesmo, um polipeptídeo maduro, ou um polipeptídeo de comprimento completo do polipeptídeo tendo atividade biológica.

28. Proteína de fusão de acordo com qualquer uma das reivindicações 20-27, caracterizada pelo fato de ainda compreender um ligador localizado entre pelo menos o domínio catalítico de endoglucanase ou a porção do mesmo e pelo menos o domínio catalítico do polipeptídeo tendo atividade biológica ou a porção do mesmo e alternativamente também um sítio de clivagem localizado entre pelo menos o domínio catalítico de endoglucanase ou a porção do mesmo e pelo menos o domínio catalítico de um polipeptídeo tendo atividade biológica ou a porção do mesmo.

29. Proteína de fusão de acordo com qualquer uma das

reivindicações 20-28, caracterizada pelo fato de ainda compreender uma quarta sequência de aminoácido compreendendo um segundo peptídeo de sinal.

5 30. Proteína de fusão de acordo com qualquer uma das reivindicações 20-29, caracterizada pelo fato de que o segundo peptídeo de sinal é ligado em armação ao N-terminal da sequência de aminoácido compreendendo pelo menos o domínio catalítico do polipeptídeo tendo atividade biológica ou a porção do mesmo ou pelo menos o domínio catalítico da endoglucanase ou a porção do mesmo.

10 31. Proteína de fusão de acordo com qualquer uma das reivindicações 20-30, caracterizada pelo fato de que o polipeptídeo tendo atividade biológica é um anticorpo, antígeno, peptídeo antimicrobiano, enzima, fator de crescimento, hormônio, imunodilatador, neurotransmissor, receptor, proteína de repórter, ou proteína estrutural.

15 32. Proteína de fusão de acordo com qualquer uma das reivindicações 20-31, caracterizada pelo fato de que o polipeptídeo tendo atividade biológica é uma oxidoreductase, transferase, hidrolase, liase, isomerase ou ligase.

20 33. Proteína de fusão de acordo com qualquer uma das reivindicações 20-32, caracterizada pelo fato de que o polipeptídeo tendo atividade biológica é uma enzima celulolítica ou uma enzima hemicelulolítica.

34. Proteína de fusão de acordo com a reivindicação 33, caracterizada pelo fato de que a enzima celulolítica é uma beta-glucosidase.

25 35. Proteína de fusão de acordo com a reivindicação 34, caracterizada pelo fato de que a beta-glucosidase compreende uma sequência de aminoácido tendo pelo menos 70%, preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 85%, o mais preferivelmente pelo menos 90%, e ainda mais preferivelmente pelo menos 95% de identidade para SEQ ID NO: 24 ou uma beta-glucosidase

variante compreendendo uma sequência de aminoácido tendo pelo menos 70%, preferivelmente pelo menos 75%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 85%, o mais preferivelmente pelo menos 90%, e mesmo o mais preferivelmente pelo menos 95% de identidade para SEQ ID NO: 35.

36. Proteína de fusão de acordo com a reivindicação 34, caracterizada pelo fato de que a beta-glucosidase é SEQ ID NO: 24 ou SEQ ID NO: 35.

37. Proteína de fusão de acordo com qualquer uma das reivindicações 20-36, caracterizada pelo fato de que é uma proteína de fusão de beta-glucosidase ou componentes da mesma.

38. Proteína de fusão de acordo com a reivindicação 37, caracterizada pelo fato de que compreende SEQ ID NO: 74 ou SEQ ID NO: 76.

39. Proteína de fusão de acordo com a reivindicação 38, caracterizada pelo fato de que SEQ ID NO: 74 é codificada por SEQ ID NO: 73 e SEQ ID NO: 76 é codificada por SEQ ID NO: 75.

40. Proteína de fusão de acordo com qualquer uma das reivindicações 20-40, caracterizada pelo fato de que a proteína de fusão, um componente a mesma, ou uma combinação da mesma, tem atividade celulolítica ou atividade hemicelulolítica.

41. Polinucleotídeo isolado, caracterizado pelo fato de codificar a proteína de fusão como definida em qualquer uma das reivindicações 20-40.

42. Construção de proteína de fusão, caracterizada pelo fato de compreender o polinucleotídeo definido na reivindicação 41.

43. Vetor de expressão, caracterizado pelo fato de compreender a construção de proteína de fusão definida na reivindicação 42.

44. Célula hospedeira fúngica, caracterizada pelo fato de ser

transformada com a construção de proteína de fusão definida na reivindicação 43.

45. Método para degradar ou converter um material celulósico, caracterizado pelo fato de compreender: tratar o material celulósico com uma quantidade eficaz de uma composição de enzima celulolítica na presença de uma quantidade eficaz da proteína de fusão da reivindicação 40 ou um componente da mesma.

46. Método de acordo com a reivindicação 45, caracterizado pelo fato de que a composição de enzima celulolítica compreende uma ou mais proteínas celulolíticas que são selecionadas dentre o grupo consistindo de uma celulase, uma endoglucanase, uma celobiohidrolase, e um polipeptídeo tendo atividade de melhora celulolítica.

47. Método de acordo com a reivindicação 45 ou 46, caracterizado pelo fato de ainda compreender tratar o material celulósico com uma quantidade eficaz de uma ou mais enzimas selecionadas dentre o grupo consistindo de uma hemicelulase, esterase, protease, laccase, peroxidase, ou uma mistura das mesmas.

48. Método de acordo com qualquer uma das reivindicações 45-47, caracterizado pelo fato de que o método é um processo de pré-tratamento, uma etapa em um processo de sacarificação e fermentação simultâneas (SSF), uma etapa em uma hidrólise e fermentação separadas (SHF) ou uma etapa em um processo de hidrólise e fermentação híbrido (HHF).

49. Método de acordo com qualquer uma das reivindicações 45-48, caracterizado pelo fato de ainda compreender recuperar o material celulósico degradado.

50. Método de acordo com a reivindicação 45, caracterizado pelo fato de que a proteína celulolítica e/ou proteína de fusão ou o componente das mesmas estão na forma de um caldo de fermentação com ou

sem células.

51. Método para produzir uma substância, caracterizado pelo fato de compreender:

5 (a) sacarificar um material celulósico com uma quantidade eficaz de uma composição de enzima celulolítica na presença de uma quantidade eficaz da proteína de fusão da reivindicação 40 ou um componente da mesma;

(b) fermentar o material celulósico sacarificado da etapa (a) com um ou mais microorganismos de fermentação, e

10 (c) recuperar a substância orgânica da fermentação.

52. Método de acordo com a reivindicação 51, caracterizado pelo fato de que a composição de enzima celulolítica compreende uma ou mais proteínas que são selecionadas dentre o grupo consistindo de uma celulase, uma endoglucanase, uma celobiohidrolase, e um polipeptídeo tendo
15 atividade de melhora celulolítica.

53. Método de acordo com a reivindicação 51 ou 52, caracterizado pelo fato de ainda compreender tratar o material celulósico com uma quantidade eficaz de uma ou mais enzimas selecionadas dentre o grupo consistindo de uma hemicelulase, esterase, protease, laccase, peroxidase ou
20 uma mistura das mesmas.

54. Método de acordo com qualquer uma das reivindicações 51-53, caracterizado pelo fato de que as etapas (a) e (b) são realizadas simultaneamente em uma sacarificação e fermentação simultâneas.

55. Método de acordo com qualquer uma das reivindicações
25 51-53, caracterizado pelo fato de que a substância é um álcool, ácido orgânico, cetona, aminoácido ou gás.

56. Método de acordo com qualquer uma das reivindicações 51-55, caracterizado pelo fato de que a composição de enzima celulolítica e/ou proteína de fusão ou um componente das mesmas estão na forma de um
30 caldo de fermentação com ou sem células.

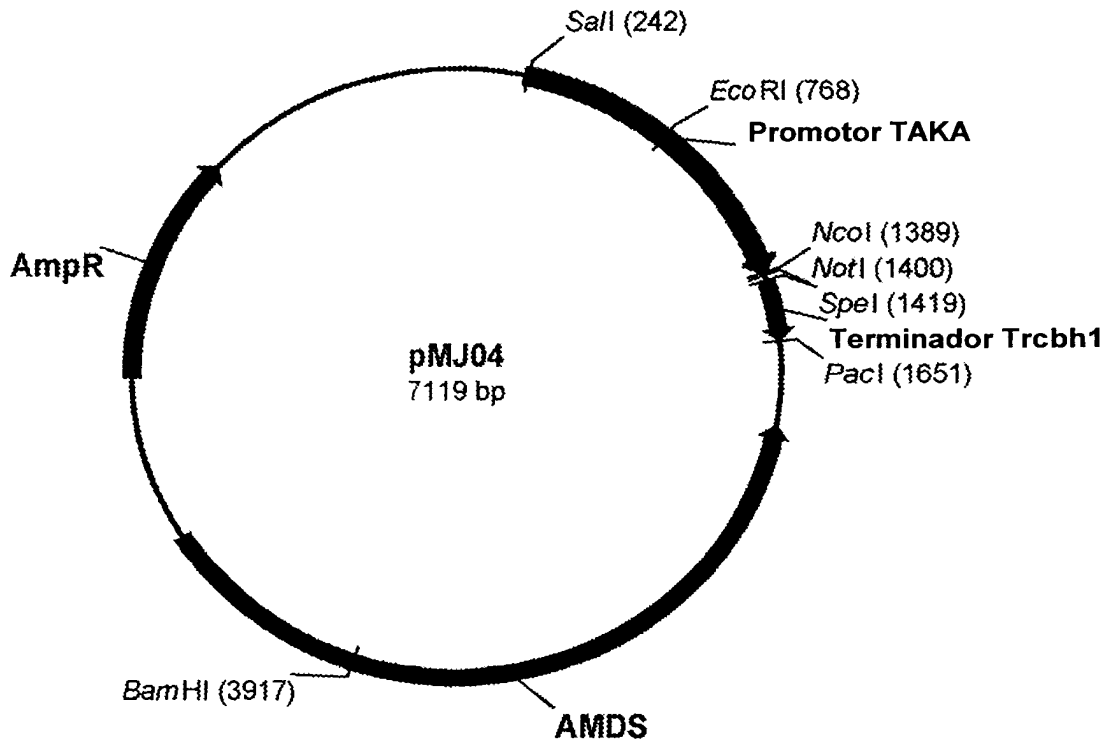


Fig. 1

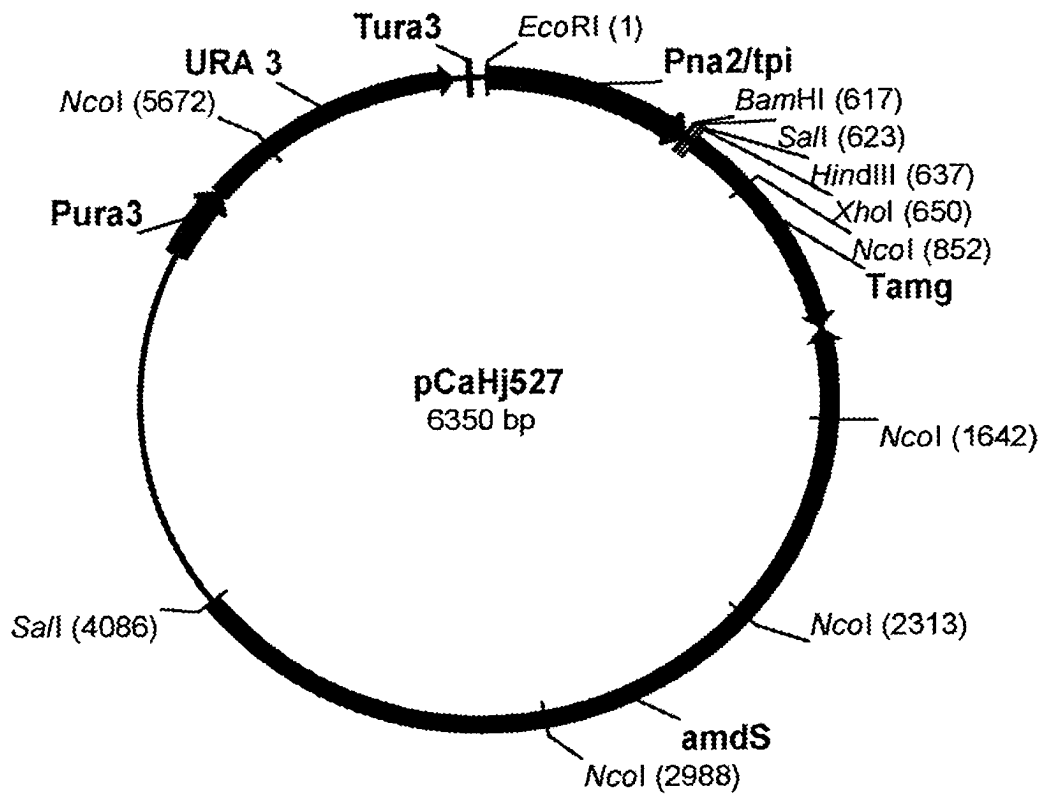


Fig. 2

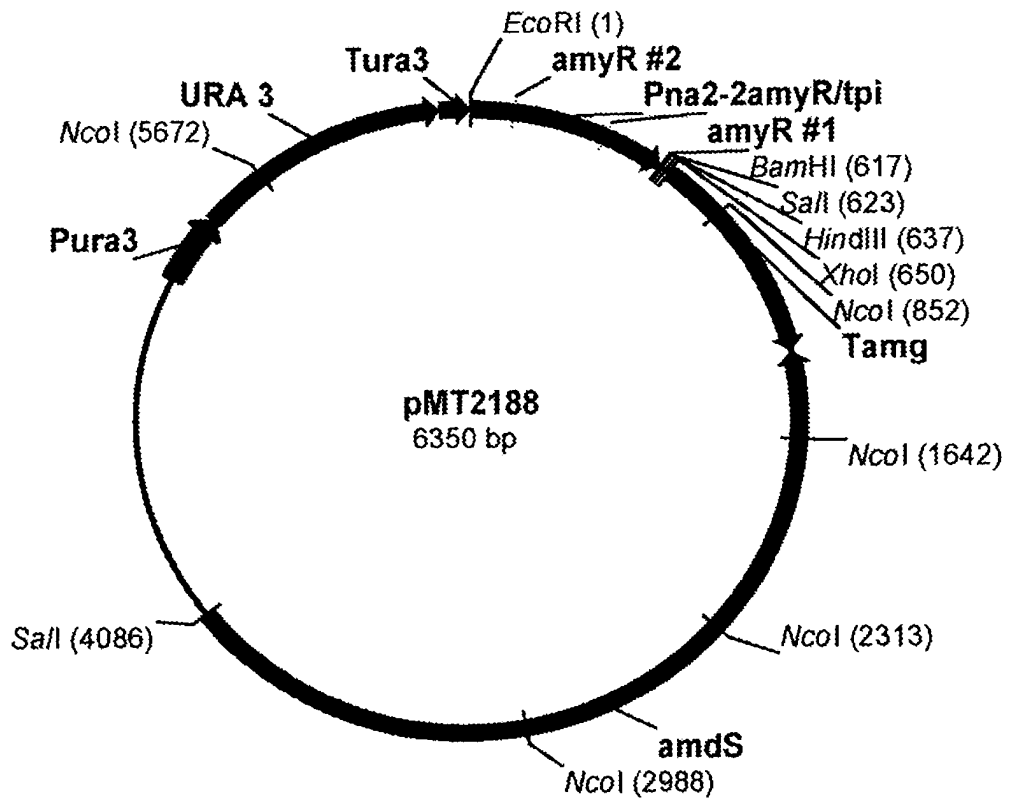


Fig. 3

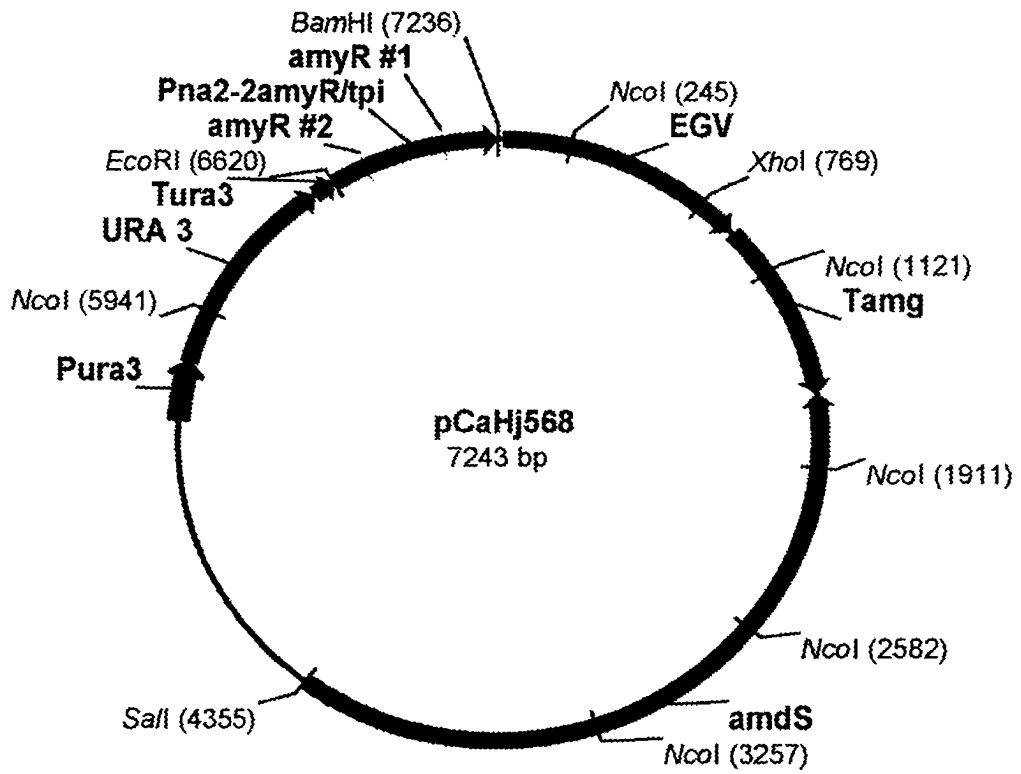


Fig. 4

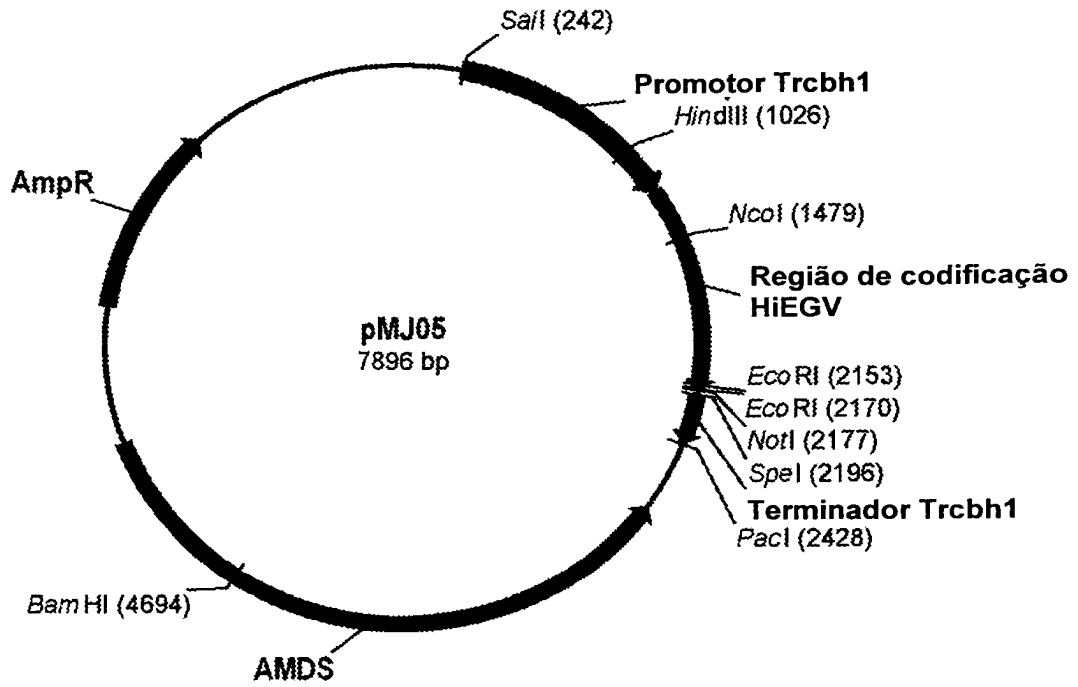


Fig. 5

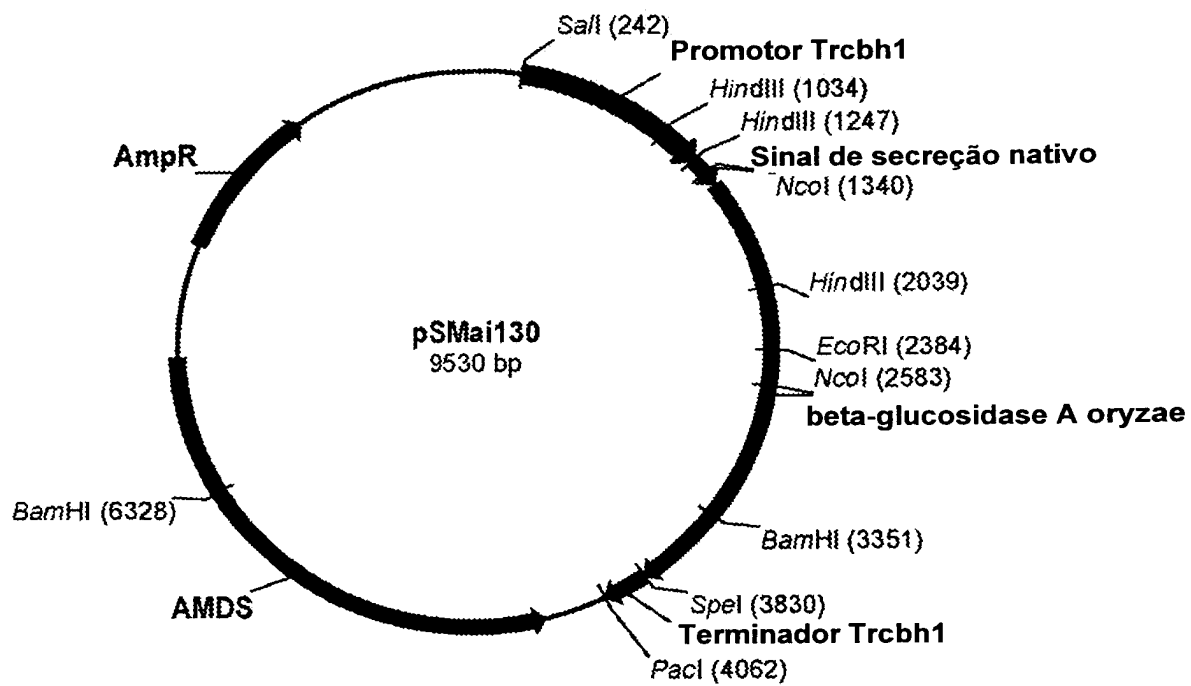


Fig. 6

ATG AAG CTT GGT TGG ATC GAG GTG GCC GCA TTG GCG GCT GCC TCA GGA GTC AGT GCC
M K L G W I E V A A L A A A S V Y S A

Fig. 7

ATG OGT TCC TCC CCC CTC CTC CGC TCC GCC GTT GTG GCC GCC CTG CCG GTG TTG GCC CTT GCC
M R S S P L L R S A V V A A L P V L A L A

Fig. 8

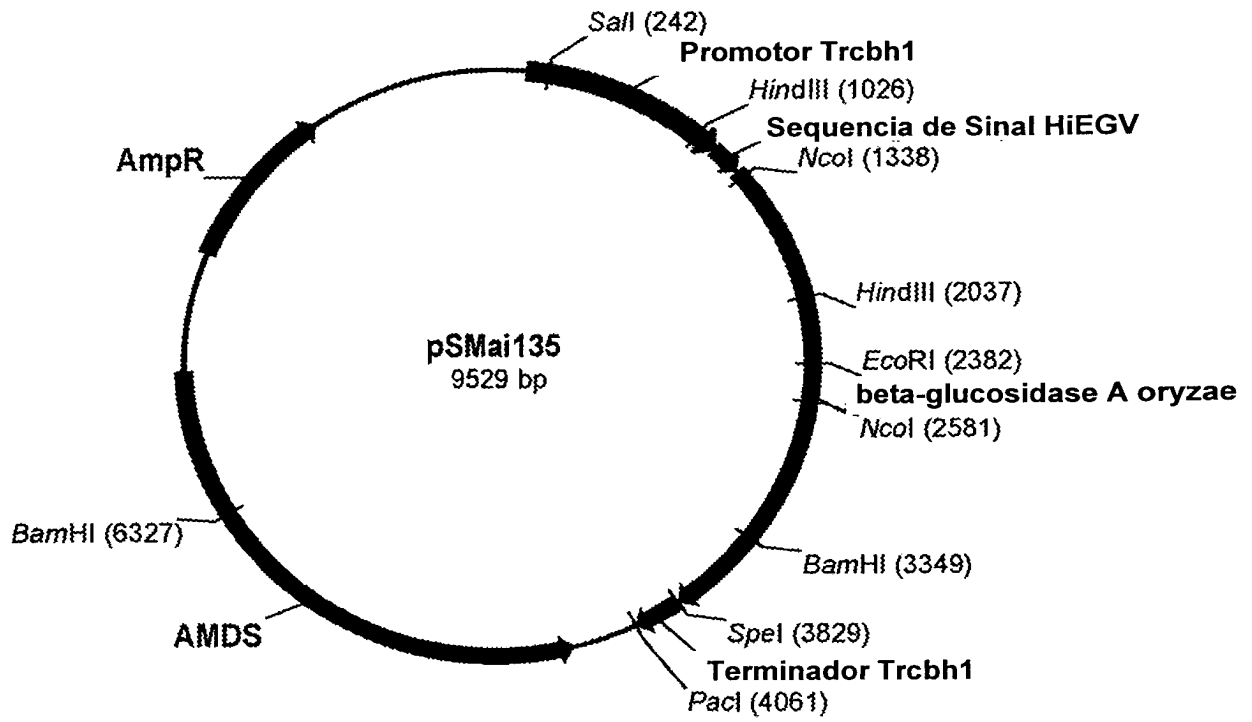


Fig. 9

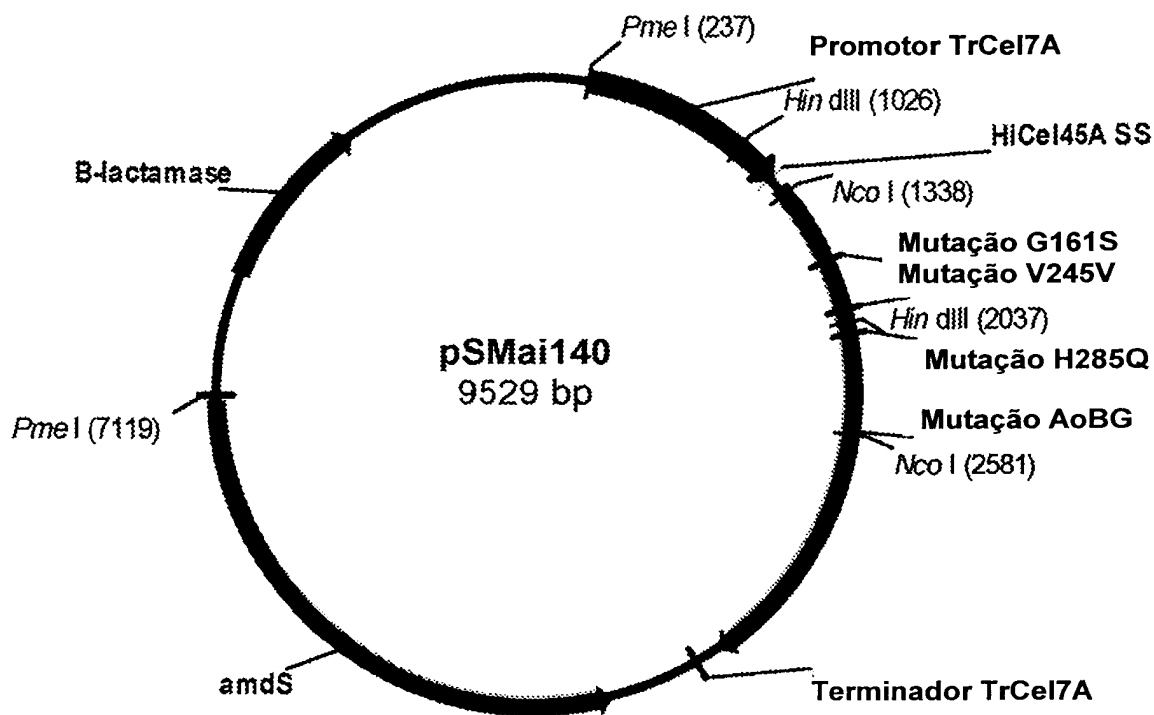


Fig. 10

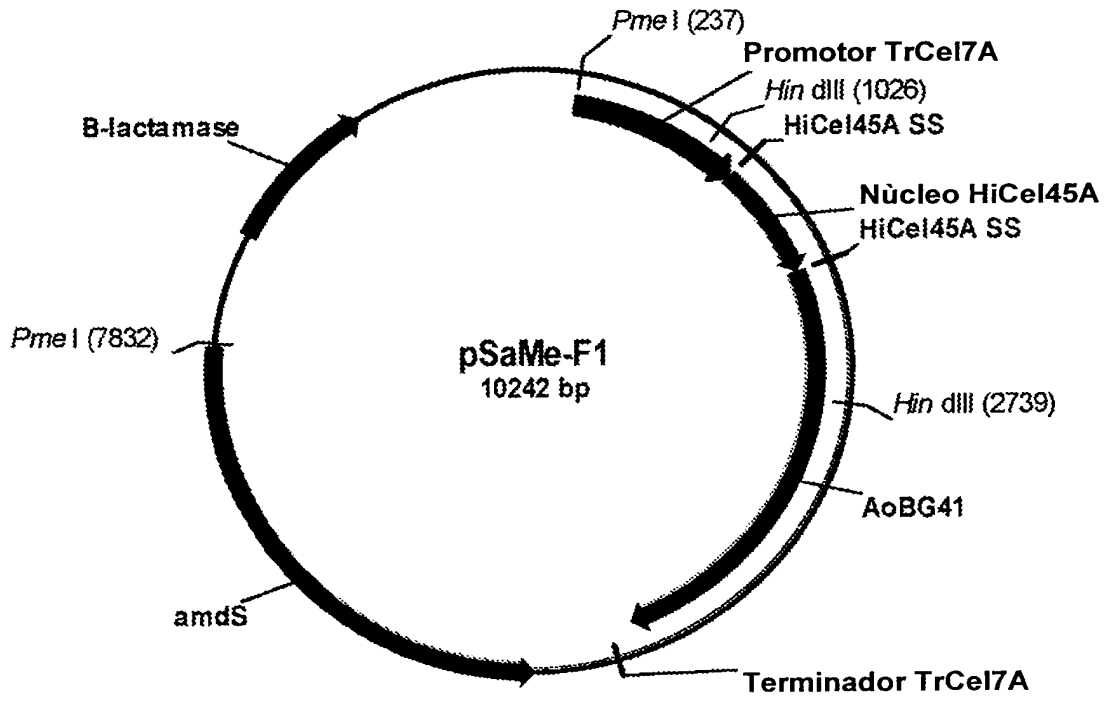


Fig. 11

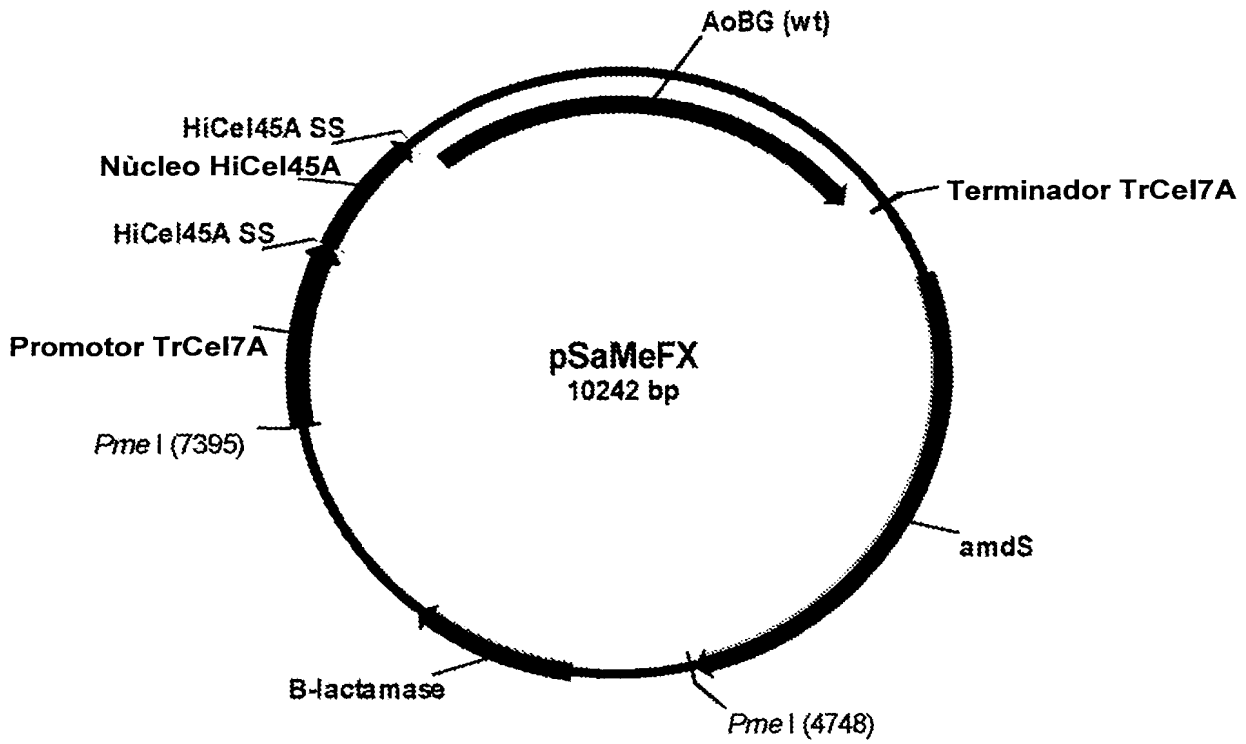


Fig. 12

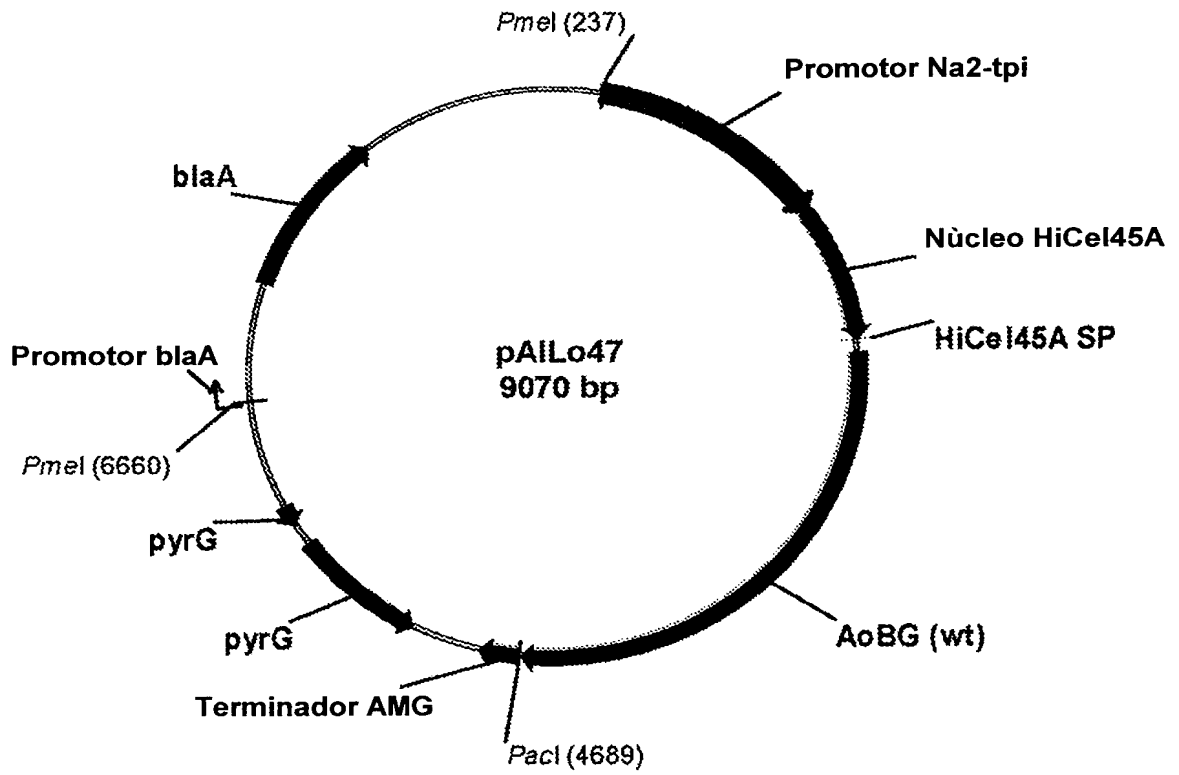


Fig. 13

M R S S P L L R S .

1291 ATGGGTTCCCTCCCCCTCCTCCGCTC
TACCACAGGAGGGGGAGGAGGCGAG
· A V V A A L P V L A L A A D G R S T R Y
1261 GCGCGTTGTGGCCGCTGCGGTGTTGGCCCTGCGCTGATGGCAGGTCCACCCGCTA
GCGGCAACACCGCGGGACGGCCACAACCGGGACCGGCGACTACCGTCCAGGTGGGCGAT
· W D C C K F S C G W A K K A P V N Q P V
1321 CTGGGACTGCTGCAAGCCTTCGTGCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGT
GACCCCTGACGACGTTCCGAAGCACGCCGACCCCGTTCCTCCGAGGGCACTTGGTCCGACA
· F S C N A N F O R I T D F D A K S G C E
1361 CTTTTCCTGCAACGCCAACTTCCAGCGTATCACGACTTCGACGCCAAGTCCGGCTGCGA
GAAAAGGACGTTGCGGTTGAAGGTCGCATAGTGCCTGAAGCTGGGTTCCAGGCCGACGCT
· P G G V A Y S C A D Q T P W A V N D D F
1441 GCGGGCGGTTGCGCTACTGCTGCGCCGACCCAGACCCCATGGGCTGTGAACGACGACT
CGGCCCGCCACAGCGGATGAGCACCGCGCTGCTGGGGTACCCGACACTTGCCTGCTAA
· A L G S F A A T S I A G S N E A G W C C A
1501 GCGGCTCGGTTTTGCTGCGCACTCTATTGCGCGGCAATGAGGGCGGCTGGTGTGCGG
GGCCGAGCCAAAGCGACGTTGGAGATAAGCGCGCTCGTTACTCCGCCCGACCAAGACCGC
· C Y E L T F F T S G P V A G K K M V V Q S
1561 CTGCTACGAGCTCACTTCAATCCGGTCTGTTGCTGGCAAGAGATGGTCTGTCAGTC
GACGATCTCGAGTGGAAAGTGTAGGCCAGGACAACGACCGTTCCTTACCAGCAGGTGAG
· T S T G D L G S N H F D L N I P G G
1621 CACCAGCACTGGCGGTGATCTTGGCAGCAACCCTTCGATCTCAACATCCCCGGCGCGG
GTGGTCTGACCGCCACTAGAACCGTCTGTTGGTGAAGCTAGAGTGTAGGGGCGCGCGCC
· V G I F D G C T P Q F G G L P G Q R Y
1661 CGTGGCATCTTCGACGGATGCACTCCCCAGTTCGGTGGTCTGCCCGCCAGCGCTACGG
GCAGCCGTAGAAGCTGCCTACGTGAGGGGTCAAGCCACCAGACGGGCGGCTCGCGATGCC
· G I S S R N E C D R F P D A L K P G C Y
1741 GGGCATCTGTTCCGCAAGAGTGGGATCGGTTCCCGACGCCCTCAAGCCCGGCTGCTA
GCCGTAGAGCAGGGCGTTCCTCACGCTAGCCAAAGGGCTGCGGAGTTCGGGCCGAGAT
· W H F D W F K N A D N P S F S F R Q V Q
1801 CTGGCGCTTCGACTGGTTCAGAAAGCCGACAAATCCGAGCTTCAGCTTCGGTCCAGTCCA
GACCGCAAGCTGACCAAGTTCCTGGCGCTGTTAGGCTCGAAGTCAAGGCCAGTCCAGGT
· C P A E L V A R T G C R R N D D G N P
1861 GTGCCCCAGCCGACTCGTCTGCGCTGCAACCGGATGCGCGCAACGACGACGGCAACTTCCC
CACGGGTCCGCTCAGCAGCGAGCGTGGCCTACGCGCGGCTTGGCTGCTGCGGTTGAAGGG
· A V Q I P M R S S P L L R S A V V A A L
1921 TGCCGTCCAGATCCCATGCGTTCCTCCCCCTCCTCCGCTCCGCGCTTGTGGCCGCTT
ACCGCAGGCTTAGGGTACGCAAGGAGGGGGAGGAGGCGAGGCGGCAACACGGCGGGA
· P V L A L A K D D L A Y S P P F Y P S P
1981 GCGGGTGTGGCCCTTGCCTAAGGATGATCTCGGTAATCCCGCTTCTTCTACCCCTCCCC
CGGCCACAACCGGAAAGGTTCTACTAGAGCGCATGAGGGGAGGAAAGATGGGAAGGGG
· W A D G Q G E W A E V Y K R A V D I V S
2041 ATGGGCGATGGTCAAGGTGATGGCGGGAAGTATACAAACCGCTGTAGACATAGTTTC
TACCCGCTTACCAGTCCCACCTTACCCGCTTCATATGTTTGGCGGACATCTGTATCAAG
· Q M T L T E K V N L T T G T G W Q L E R
2101 CCAGATGACGTTGACAGAGAAAGTCAACTTAAAGACTGGAACAGGATGGCAACTAGAGAG
GGTCTACTGCAACTGTCTCTTTCAGTTGAATTGCTGACCTTGTCTACCGTTGATCTCTC
· C V G Q T G S V P R L N I P S L C L Q D

Fig. 14A

2161 GTGNTTGGACAAACTGGCAGTGTTCACGACTCAACATCCCCAGCTTGTGTTTGCAGGA

CACACAAACCTGTTTGCACCGTACAAAGGGTCTGAGTTGTAGGGGTGGAACACAAAACGGTCTCT
· S P L G I R F S D Y N S A F P A G V N V
2221 TAGTCCCTCTTGGTATTGCTTTCTCGGACTACAATTCAGCTTTCCCTGCGGGTGTAAATGT
ATCAGGAGAACCATAGCAAAGAGCCCTGATGTTAAGTCGAAAGGGGACGCCCAATTACA
· A A T W D K T L A Y L R G Q A M G E E F
2281 CGCTGCCACCTGGGACAAGACGCTGCTACCTTCGCTGGTCAAGCAATGGGTGAGGAGTT
CGGACGGTGGACCCCTGTTCTGCGAGCGGATGGAAGCACCAGTCCGTTACCCACTCCTCAA
· S D K G I D V Q L G P A A G P L G A H P
2341 CAGTGATAAGGGTATTGACGTTCCAGCTGGGTCTGCTGCTGCTGCGCTCTCGGTGCTCATCC
GTCACCTATTTCCATAACTGCAAGTCGACCCAGGACGACCGACCCGGGAGAGCCACGAGTAGG
· D G G R N W E S F S P D P A L T G V L F
2401 GGATGGCGGTAGAACTGGGAAAGTTTCTCACCAGATCCAGCCCTCACCGGTGTACTTTT
CTTACCCCTCTTTGACCCCTTCAAAGAGTGGTCTAGGTGCGGAGTGGCCACATGAAAA
· A E T I K G I Q D A G V I A T A K H Y I
2461 TCGGAGACGATTAAGGGTATTCAAGATGCTGGTGTCAITGGACAGCTAAGCATATAT
AGCCCTCTGCTAATTCOCATAAGTTCTACGACCACAGTAACGCTGTGATTCGTAATATA
· M N E Q E H F R Q Q P E A A G Y G F N V
2521 CATGAACGAACAAGAGCATTTCCGCCAACAACCCGAGGCTGCGGGTTACCGATTCAACGT
GTACTTGGCTTGTCTCGTAAAGCGGTTGTTGGGCTCCGACGCCCAATGCCAAGTTGCA
· S D S L S S N V D D K T M H E L Y L W P
2581 AAGCGACAGTTTGGATTCCAAGGTGATGACAAGACTATGCAATGAATTGTAACCTCTGRC
TTCGCTGTCAAACCTCAAGGTTGCAACTACTGTTCTGATACGTAACATGGAGACCGG
· F A D A V R A V G A V N C S Y N Q I N
2641 CTTCGCGGATGCATACGCGCTGGAGTGGGTGCTGTATGCTGCTTTACAACCAATCAA
GAGCGCCTACGTCATGCGCGACCTCAGCCACGACAATACACGAGAATGTTGGTTAGTT
· N S Y G C E N S E T L N K L L K A E L G
2701 CAACAGCTACGGTTCGAGATAGCGAAACTCTGAACAAGCTTTTGAAGCGGAGCTTGG
GTTGTGATGCCAACGCTCTTATGCTTTGAGACTTGTTCGAAAACCTCCGCTCGAACC
· F Q G F V M S D W T A Q H S G V G A A L
2761 TTTCCAAGGCTTCGTCATGAGTGTGGACCGCTCAACACAGCGCGGTAGGCGCTGCTTT
AAAGGTTCCGAGCAGTACTCACAACCTGGCGAGTTGTTGCTCGCGCATTCCGCGAAGAA
· A G L D M S M P G D V T F D S G T S F W
2821 AGCAGGCTCTGGATATGTCGATGCGCGGTGATGTTACCTTCGATAGTGTACGCTTTCTG
TCCTCCAGACCTATACAGCTACGGGCCACTACAATGGAAGCTATCACCATGCAGAAAGAC
· G A N L T V G V L N G T I P Q W R V D D
2881 GGGTGCAAACTTGACGGTCCGCTGCTTAAACGGTACAATCCGCCAATGGCGTGTGATGA
CCCAGTTTGAAGTGGCAGCCACAGGAATTGCCATGTTAGGGGTTACCGCACAACTACT
· M A V R I M A A Y Y K V G R D T K Y T P
2941 CATGGCTGTCCGATCATGCGCGCTTATTACAAGGTTGGCCGCGACACCAAAATACACCCC
GTACCCGACAGGCATAGTACCGGCAATAATGTTCCAACCGGCGCTGTGGTTATGTGGGG
· P N F S S W T R D E Y G F A H N H V S E
3001 TCCCAACTTCAGCTCGTGGACCAGGACGAATATGGTTCGCGGCATAACCATGTTTCGGA
AGGTTGAAGTCCAGCACCTGGTCCCTGCTTATACCAAAGCGCGTATTGGTACAAAGCCT
· G A Y E R V N E F V D V Q R D H A D L I
3061 AGGTGCTTACGAGAGGGTCAACGAATTCGTGGACGTCGAAGCGATCATGCGGACCTAAT
TCCAGGAATGCTCTCCAGTTGCTTAAGCACCTGCACGTTGCGCTAGTACGGCTGGATTA
· T A N F P Y L V T P E Q A I Q N E V L Q
3121 CCGTGCATCGCGCGCAGAGCACTGTTCTGCTGAAGAACAAGGGTGCCTTGCCCTTGAG
GGCAGCGTAGCCCGCGCTCTGTGACAAGACGACTTCTTCTTCCACCGAAGCGGAACTC
· R K R K L V A L L G S D A G S N S W G A

Fig. 14B

3181 CCGCAAGGAAAAGCTGGTCCGCCCTTCTGAGAGAGGATGCGGGTTCGAACTGGTGGGGCGC
GGCGTTCCCTTTGACACCGGGAAGACCCTTCTACGCCCAAGGTTGAGCACCCCGCG
· N G C D D R G C D N G T L A M A W G S G
3241 TAACGGCTGTGATGACCGTGGTGGGATAACGCTACCCCTTGCATGCGCTGGGGTAGCGG
ATTGCGGACACTACTGGCACCAACGCTATTGCCATGGGAACGGTACCGGACCCCATCGCC
· T A N F P Y L V T P E Q A I Q N E V L Q
3301 TACTGCGAATTTCCATACTCGTGACACAGAGCAGGCGATTTCAGAACGAAGTTCTTCA
ATGAGGCTTAAAGGGTATGGACCACTGTGCTCTCGTCCGCTAAGTCTTGGCTTCAAGAGT

· G R G N V F A V T D S W A L D K I A A A ·
 3361 GGGCCGTTGGTAATGTCTTCCCGGTGACCGACAGTTEGGCCGCTCGACAAGATCGCTGCGCGC
 CCGCGCACCATACAGAAGCGGCACCTGCTGTCAACCGCGAGCTGTTCTAGCGACGCGG
 · A R Q A S V S L V F V N S D S G E G Y L ·
 3421 TGCCCGCCAGGCCAGCGTATCTCTCTGTTTCGTCAACTCCGACTCAGGAGAGGGCTATCT
 ACGGGCGGTCCGGTCCCATAGAGAGCACAAAGCAGTTGAGGCTGAGTCCCTTTCGGATAGA
 · S V D G N E G D R N N I T L W K N G D N ·
 3481 TAGTGTGGATGGAAATGAGGGCGATCGTAACAACATCACTCTGTGGAAGAACGGCGACAA
 ATCACACCTACCTTTACTCCCGCTAGCATTGTTGTAGTGAGACACCTTCTTCCCGCTGTT
 · V V K T A A N N C N N T V V I I H S V G ·
 3541 TGTGGTCAAGACCGCAGCGAATAACTGTAACAACACCGTTGTCTATCATCCACTCCGCTGG
 ACACCACTTCTGGCGTCCGCTTATGACATTTGTTGCGCAACAGTAGTAGGTTGAGGCAGCC
 · P V L I D E W Y D H P N V T G I L W A G ·
 3601 ACCAGTTTTGTATCGATGAATGGTATGACCACCCCAATGTCACTGGTATTCTCTGGGCTGG
 TGGTCAAAAAGTACTTACCATACTGGTGGGGTTACAGTGACCATAAGAGACCCCGACC
 · L P G Q E S G N S I A D V L Y G R V N P ·
 3661 TCTGCCAGGUCAGGACTCTGGTAACTCCATTGCGGATGTGCTGTACGGTCTGTCAACCC
 AGACGGTCCGGTCCCTAGACCATTGAGGTAACGGCTACACGACATGCCAGCACAGTTGGG
 · G A K S P F T W G K T R E S Y G S P L V ·
 3721 TGGGGCCAAAGTCTCCCTTTCACCTTGGGGCAAGACCCGGAGTCCGTATGGTTCTCCCTGGT
 ACCCGCGTTCAGAGGAAAGTGAACCCCGTTCTGGGCCCTCAGCATACCAAGAGGGAAACCA
 · K D A N N G N G A P Q S D F T Q G V F I ·
 3781 CAAGGATGCCAACAAATGGCAACGGAGCCCGCAGTCTGATTTTCAACCCAGGGTNTTTTCAT
 GTTCTACGGTTGTTACCGTTGCCCTCGCGGGGTCAGACTAAAGTGGGTCCACAAAAGTA
 · D Y R H F D K F N E T P T Y E F G Y G L ·
 3841 CGATTAACCGCCATTTCCGATAAGTTCAATGAGACCCCTATCTACGAGTTTGGCTACGGCTT
 CCTAATGGCGGTAAGCTATTCAAGTTACTCTGGGATAGATGCTCAAACCGATGCCGAA
 · S Y T T P E L S D L H V Q P L N A S R Y ·
 3901 GAGCTACACCACCTTCGAGCTCTCCGACCTCCATGTTTCAGCCCCGTAACCGGTCCCGATA
 CTGGATGTGGTGGAAAGCTCGAGAGGCTGGAGGTACAAGTCCGGGACTTGGCGAGGGCTAT
 · T F T S G M T E A A K N F G E I G D A S ·
 3961 CACTCCCACCACTGCCATCACTGAAGCTGCAAAAGAACTTTGGTGAATTTGGCGATGCGTC
 GTGAGGGTGGTCAACCTACTGACTTTCGACGTTTCTTGAACCACTTTAACCGCTACGCAG
 · E Y V Y P E G L E R I H E F I Y P W I N ·
 4021 GGAGTACGTGTATCCGGAGGGGCTGGAAAGGATCCATGAGTTTATCTATCCCTGGATCAA
 CCTCATGCACATAGCCCTCCCGGACCTTTCCCTAGGTAATCAAATAGATAGGGACCTAGTT
 · S T D L K A S S D D S N Y G W E D S K Y ·
 4081 CTCTACCGACCTGAAGGATCGTCTGACGATTCTAACTACGGCTGGGAGACTCCAAGTA
 GAGATGGCTGGACTTCCGTAGCAGACTGCTAAGATTGATGCCGACCCCTCTGAGGTTCTAT
 · I P E G A T D G S A Q P R L P A S G G A ·

Fig. 14C

4141 TATTCCCGAAGGGCCACGGGATGGGTCTGCCCGAGCCCGTFFGCCGGCTAGTGGTGGTGC
 ATAAGGGCTTECGGGGTGCCTACCCAGACGGGTGGGGCAAACGGGGCGATCADCACCACG
 · G G N P G L Y E D L F R V S V K V K N T
 4201 CGGAGGAAACCCCGGTCTGTACGAGGATCTTTTCCGGCTCTCTGTGAAGGTCAAGAACAC
 GCCTCTTTTGGGGCCAGACATGCTCCTAGAAAAGGCGCAGAGACACTTCCAGTTCCTTGTG
 · G N V A G D E V P Q L V V S L G G P N E
 4261 GGGCAATGTCCGGGTGATGAAGTTCCTCAGCTGTACGTTTCCCTAGGGCGGCCOGAATGA
 CCGTTACAGCGGCCACTACTTCAAGGAGTCCACATGCAAGGGATCCGCCGGGCTTACT
 · P K V V L R K F E R I H L A P S Q E A V
 4321 GCCCAAGGTGGTACTGCGCAAGTTTGAGCGTATTCACTTGGCCCEFTCGGAGGAGGCCGT
 CGGTTCCACCATGACCGGTTCAAACTCCGATAAGTGAACCGGGGAAGGCTCCTCCGKCA
 · W T T T L T R R D L A N W D V S A Q D W
 4381 GTGGACAAAGACCCCTTACCCGTCCGTGACCTTGCAAACTGGGAAGTTTCCGCTCAGGACTG
 CACCGTTGCTGGGAATGGGCAGCACGGAACGTTTGACCCCTGCAAAGCCGAGTCCCTGAC
 · T V T P Y P K T I Y V G N S S R K L P L
 4441 GACCGTACTCCTTACCCCAAGACGATCTACGTTGAAACTCCTTACGAAACTGCGCGCT
 CTGGCAGTGAAGAAATGGGGTTCGCTAGATGCAACCTTTGAGGAGTGCCTTTGACGGCGA
 · Q A S L P K A Q *
 4501 CCAGGCTCGCTGCCCTAAGGCCAGTAA
 GGTCCGGAGCGACGGATTCCGGGTCAAT

Fig. 14D

M R S S P L L R S

1201 ATGGCGTTCCCTCCCCCTCCTCCCGCTC
TACGCAAGGAGGGGGGAGGAGGCGAG
· A V V A A L P V L A L A A D G R S T R Y ·
1261 GCGCGTTGTGGCCGCTGCGCGTGTGGCCCTTGGCCCTGATGGCAGGTCCACCCGCTA
GUGGCAACACCGGCGGACCGGCCACCAACUGGGAACCGGGACTACCGTCCAGGTGGGCGAT
· W D C C K P S C G W A K K A F V N Q P V ·
1321 CTGGGACTGCTGCAAGCCTTCGTGCGGCTGGGGCCAAGAAGGCTCCCGTGAACCAGCCTGT
GACCCGTGACGAGTTCGGAAGCACGCCGACCCGTTCTTCGAGGGCACTTGGTGGACA
· F S C N A N F Q R I T D F D A K S G C E ·
1381 CTFTTCCTGCAACGCCAACTTCCAGCGTATCACGGACTTCGACGCCAAGTCCGGCTGGCA
GAAAAGGACGTTGCGGTTGAAGGTGCGATAGTCCCTGAAAGCTGCGGTTTCAGGCCGACCGT
· P G G V A V S C A D Q T P W A V N D D F ·
1441 GCGGGGCGGTGTGCGCTACTGCTGCGCCGACCCAGACCCATGGGCTGTGAACGACGACTT
CGCCCCGCCACAGCGGATGAGCACCGGCTGGTCTGGGTTACCCGACACTTGCCTGCTGAA
· A L G F A A T S I A S S N E A G W C C A ·
1501 CGCGCTCGGTTTTGCTGCCACCTCTATTTGCCGGCAGCAATGAGGGGGGCTGGTGTGCGC
GCGCGAGCCAAAACGACGGTGGAGATAACGGCCGTCGTTACTCCGCCGACCCAGCAACCGG
· C Y E L T F T S G P V A G K K M V V Q S ·
1561 CTGCTACGAGCTCACCTTCACATCCGCTCCTGTTGGCTGGCAAGAAGATGGTGTGCTCAGTC
GACGATGCTCGAGTGAAGTGTAGGCCAGGACAACGACCGTTCTTCTACCCAGCAGSTCAG
· T S T G G D L G S N H F D L N I P G G G ·
1621 CACCAGCACTGGCGGTGATCTTGGCAGCAACCACCTTCGATCTCAACATCCCGGGCGGGCG
GTGGTCTGTGACCGCCACTAGAACCGTCTGGTGGTGAAGCTAGAGTTGTAGGGGCGCCGCC
· V G I F D G C T P Q F G G L P G Q R Y G ·
1681 CGTCGGCATCTTCGACCGGATGCACCTCCCAGTTCGGTGGTCTGCCCGGCCAGCGCTACGG
GCAGCCGTAGAAGCTGCCTACGTGAGGGGTCAAGCCACCAGACGGGCGCGTCCGGATGCC
· G I S S R N E C D R F P D A L K P G C Y ·
1741 CGGCATCTCGTCCCGCAACGAGTGGGATCGGTTCCCGGACGCCCTCAAGCCCGGCTGCTA
GCCGTAGAGCAGGGCGTTGCTCACGCTAGCCCAAGGGGCTGCGGGAGTTCGGGCGCAGCAT
· W R F D W F K N A D N P S F S F R Q V Q ·
1801 CTGGCGCTTCGACTGGTTCAGAAACCGCGACAATCCGAGCTTCAGCTTCGGTCCAGGTCCA
GCCGCGAAGCTGACCAAGTTCCTTGGCGCTGTTAGGTTCGAAGTCGAAGGCAGTCCAGGT
· C P A E L V A R T G C R R N D D G N F P ·
1861 GTGCCCAGCCGAGCTCGTCTGCTCGCACCGGATGCGCGCAACGACGACGGCAACTTCCC
CACGGTTCGGCTCGAGCAGCGAGCGTGGCCCTACGGCGCGGTTGCTGCTGCGGTTGAAGGG
· A V Q I P M R S S P L L R S A V V A A L ·
1921 TGCCGTTCCAGATCCCCATGCGTTCCTCCCCCTCTCTCGCTCCGCGCTTGTGGCCGCCCT
ACGGCAGGTCTAGGGGTACGCAAGGAGGGGGAGGAGGCGAGGCGGCAACACCGGCGGGA
· P V L A L A K D L A Y S P P F Y P S P ·
1981 GCGCGTGTGGCCCTTSCCAAGGATGATCTCGGTACTCCCTCTTTTCTACCCCTTCCCC
CGGCCACAACCGGAACGGTTCCTACTAGAGCCCATGAGGGGAGGAAAGATGGGAAGGGG
· W A D G Q G E W A E V Y K R A V D I V S ·
2041 ATGGGCAGATGGTTCAGGGTGAATGGGCGGAAGTATACAAAACGGGCTGTAGACATAGTTTC
TACCGCTTACCACTCCACTTACCCGCCCTTCATATGTTTGGCGACATCTGTATCAAAG
· Q M T L T E K V N L T T G T G W Q L E R ·

Fig. 15A

2101 CCAGATGACGTTGACAGAGAAAGTCAACTTAACGACTGGAAACAGGATGGCAACTAGAGAG
GGTCTACTGCAACTGTCTCTTTCAGTTGAATTGCTGACCTTGTCTACCGTTGATCTCTC
· C V G Q T G S V P R L N I P S L C L Q D ·
2161 GGTGTGTGGACAAAACGTGGCAGTGTTCACGACTCAACATCCCGAGCTTGTGTTTGGCAGGA

CACACAACCTGTTTGACCGTCAAGGGTCTGAGTTGTAGGGTTCGAACACAAACGTCCT
 · S P L G I R F S D Y N S A F P A G V N V
 2221 TAGTCCTCTTGGTATTGGTTTCTCGGACTACAATTCAGCTTTCCCTGCGGGTGTAAATGT
 ATCAGGAGAACCATAAGCAAAGAGCCTTGATGTTAAGTTCGAAAGGGACGCCACAAATTACA
 · A A T W D K T L A Y L R G Q A M G E E F
 2281 CGCTGGCCACCTGGGACAAGACGGTGGCCTACCTTCGTGGTCAGGCAATGGGTGAGGAGTT
 GCGAGGGTGGACCCCTGTTCTGCGAGCGGATCGAAGCACCAGTCCGTTACCCACTCCCTCAA
 · S D K G I D V Q L G P A A G P L G A H P
 2341 CAGTGATAAGGGTATTGACGTTCCGCTGGGTCCGCTGCTGGCCCTCTCGGTGCTCATCC
 GTCACTATTCCCATAACTGCAACTCGACCCAGGACGACGACCCGGAGAGCCACGAGTAGG
 · D G G R N W E G F S P D P A L T G V L F
 2401 GGATGGCGGTAGAACTGGGAAGGTTTCTCACCAGATCCAGCCCTCACCGGTGTACTTTT
 CCTACCCCATCTTTGACCCCTCCAAAGAGTGGTCTAGGTCCGGAGTGGCCACATGAAAA
 · A E T I K G I Q D A G V I A T A K H Y I
 2461 TCGGAGACGATTAAGGTTATTCAAGATGCTGGTGTCTATGCGGACAGCTAAGCAATTATAT
 ACGCCTCTGCTAATTCOCATAAGTTCTACGACCACAGTAACGCTGTCGATTTCGTAATAA
 · M N E Q E H F R Q Q P E A A G Y G F N V
 2521 CATGAACGAACAGAGCATTTCGCCAACACCCGAGGCTGCGGCTTACGGATTCAACGT
 GTACTTGTCTTCTCGTAAAGGGGTTGTTGGGCTCCGACGCCCAATGCTTAAGTTGCA
 · S D S L S S N V L D K T M H E L Y L W P
 2581 AAGCGACAGTTTGGATTCCAACGTTGATGACAAGACTATGCATGAATTGTACCTCGGGC
 TTGCTGTCAAACTCAAGGTTCAACTACTGTTTGTATCGTACTTAACATGGAGACCCG
 · F A D A V R A G A V M C S Y N Q I N
 2641 CTTGCGGGATGCAGTACCGCGCTGAGTCCGCTGCTGTCTGCTCTTACAACCAATCAA
 GAAGCGCCTACGTCATGCGCGACTCAGCCACGACAGTACCGAGAATGTTGGTTTACTT
 · N S Y G C E N S E T L N K L L K A E L G
 2701 CAACAGCTACGGTTGCGAGAATAGCGAAACTCTGAACAAGCTTTTGAAGGCGGAGCTTGG
 GTTGTGATGCCAACGCTCTTATCGCTTTGAGACTTGTTCGAAACTTCCGCCPCGAACC
 · F Q G F V M S D W T A H H S G V G A A L
 2761 TTTCCAAGGTTCTCATGAGTATGGACCGCTCATCACAGCGGCTTAGGGCTGCTTT
 AAAGGTTCCGAGCAGTACTCCTAACCTGGCGACTAGTNTCCGGCATCCCGGACGAAA
 · A G L D M E M P G D V T F D S G T S F W
 2821 AGCAGGTCGGATATGTCGATGCCCGGTGATGTTACCTTCGATAGTGGTACGCTTTCTG
 TCGTCCAGACCTATACAGCTACGGGCCACTACAAATGGAAGCTATCACCATGCAGAAAGAC
 · G G A N L T V G V L N G T I P Q W R V D D
 2881 GGGTGCAAACTTGACCGTGGTCTCCCTTAAAGGTACAATCCCCCAATGGCGTNTTGAAGA
 CCCACGTTTGAACCTGCCAGCCACAGGAATTGCCATGTTAGGGGTTACCGCACCACTACT
 · M A V R E M A A Y Y K V G R D T K Y T P
 2941 CATGGCTGTTCCGTATCATGGCCGCTTATTACAAGGTTGGCCGCGCACCAAAATACACCCC
 GPACCGACAGGCATAGTACCGGAAATAATGTTCCAACCGCGCTGTGGTTTATGTGGGG
 · P N F S S W T R D E Y G F A H N H V S E
 3001 TCCCAACTTCAGCTCGTGGACCAGGACGAATATGGTTTCCGGCATAACCATGTTTCCGA
 AGGTTGAAGTGCAGCACCTGCTCCCTTATACCAAGCGCTATTGGTACAAAGCCT
 · G A N F R V N E F V D V Q R D H A D L I
 3061 AGGTGCTTACGAGAGGGTCAACGAATTCGTTGACCGTCAACCGGATCATGCCGACCTAAT
 TCCAGGAATGCTCTCCAGTTGCTTAAAGCACCCTGCAAGTGTGGCTAGTACGGTGGATTA
 · R R I G A Q S T V L L K N K G A L P L S

Fig. 15B

3121 CCTTCGCATCGGCGCGCAGAGCACTGTTCTGCTGAAGAACAAGGGTCCCTTCCCTTGGAG
 UGCAGCGTAGCCGCGCTCTCGTGACAAGACCACTTCTTGTTCCLACGGAACGGGAACTC
 · R K E K L V A L L G E D A G S N S W G A
 3181 CCGCAAGGAAAAGCTGCTCGECCTTETGGGAGAGGATGCGGGTCCACTCGTGGGGCGC
 GGCGTTCCTTTTUGACCGAGCGGGAAGACCTCTCTACGCCCAAGGTTGAGCACCCCGCG
 · N G C D D R G C D N G T L A M A W G S G
 3241 TAACGGCTGTGATGACCGTGGTTGCGATAACGGTACCCTTGCATGGCCTGGGGTAGCGG
 ATTGGCCGACACTACTGGCACCAACGCTATTTGCCATGGGAACGGTACCGGACCCCATCGCC
 · T A N F P Y L V T F E Q A I Q N E V L Q
 3301 TACTGCCAAATTTCCCATACCTCGTGAACACAGAGCAGGCGATTCAGAAACGAAGTCTTCA
 ARGACGCTTAAGGGTATGAGGACCTGTTGGTCTCGTTCGGCTAAGTCTTGGCTTCAAGAAGT

· G R G N V F A V T D S W A L D K I A A A ·
 3361 GGGCCGTTGTAATGTCTTCGCGGTGACCGACAGTTGGGGCGTTCGACAAGATCGCTGCGGC
 CCCGGCACCATTACAGAAGCGGCACCTGCGCTGTCAACCCGCGAGCTGTTCTAGCGACGCCG
 · A R Q A S V S L V F V N S D S G E G Y L ·
 3421 TGCCCGCCAGGCCAGCGTATCTCTCGTGTTCGTCAACTCCGACTCAGGAGAAGGCTATCT
 ACCGGCGCTCCGCTCGCATAGAGAGCACAAAGCAGTTGAGGCTGAGTCCCTCTTCGGATAGA
 · S V D G N E G D R N N I T L W K N G D N ·
 3481 TAGTGTGGATGGAAATGAGCGGCATCGTAACAACATCACTCTGTGGAAGAACGGCGGACAA
 ATCACACCTACCTTTACTCCGCTAGCATTTGTTGTAGTGAGACACCTTCTTCCGCTGTT
 · V V K T A A N N C N N T V V I I H S V G ·
 3541 TGTGGTCAAGACCGCAGCGAATAACTGTAACAACACCGTTTTCATCATCCACTCCGTCGG
 ACACCAGTTCTGCGCTCCCTTATTGACATTTGTTGTGGCAACAGTAGTAGGTTGAGGCAGCC
 · P V L I D E W Y D H P N V T G I L W A G ·
 3601 ACCAGTTTATGATCGATGAATGGTATGACCAACCCCAATGTCAGTGGTATTCTCTGGCGCTGG
 TGGTCAAACCTAGCTACTTACCATACTGGTGGGGTTACAGTGACCATAAGAGACCCGACCC
 · L P G Q E S G N S I A D V L Y G R V N P ·
 3661 TCTGECAGGCCAGGAGTCTGGTAACCTCATTGCGGATGTCTGTACGGTCTGTCAACCC
 AGACGGTCCGGTCCCTCAGACCCATTGAGCTAACGGCTACACGACATGCCAGCACAGTTGGG
 · G A K S P F T W G K T R E S Y G S P L V ·
 3721 TGGGGCCAAGTCTCCTTTCACTTGGGCAAGACCCGGGAGTGGTATGGTTCTCCCTTGGT
 ACCGCGTTCAGAGGAAAGTGAACCCCGTTCTGGGCCCTCAGCATACCAGAGGGGAACCA
 · K D A N N G N G A P Q S E F T Q G V F I ·
 3781 CAAGGATGCCAACAAATGGCAACGGAGCGCCCGAGTCTGATTTCCACCCAGGGTGTTTTCA
 GTTCTTACGGTTGTTACCGTTGCCCTCGCGGGTTCAGACTAAAGTGGGTCCCAAAAAGTA
 · D Y R H F D K F N E T P I Y E F G Y G L ·
 3841 CGATTACCGCCATTTGATAAGTTCAATGAGACCCCTATTTACAGATTTGGCTACGGCTT
 GCTAATGGCGGTAAGCTATTCAAGTACTCTGGGGATAGATGCTCAAACCGATGCCGAA
 · S Y T T P E L S D L H V Q P L N A S R Y ·
 3901 GAGCTACACCACCTTCGAGCTCTCCGACCTCCATGTTCCAGCCCTGAAACCGCTCCCGATA
 CTCCATGTGTTGGAAAGCTCGAGAGGTTGGAAGTACAAGTCCGGGACTTGGCCAGGGCTAT
 · T P T S G M T E A A K N F G E I G D A S ·
 3961 CACTCCACCAGTGGCATGACTGAAGCTGCAAGAAGCTTTGGTGAATTTGGCGATGCGTC
 GTGAGGGTGGTCAACCTACTGACTCGAAGTTTCTTGAACCACTTTAACCGCTACGCGAG
 · E Y V Y P E G L E R I H E F I Y P W I N ·

Fig. 15C

```

4021 GGAGTACGTGTATCCGGAGGGGCTGGAAAGGATCCATGAGTTTATCTATCCCTGGATCAA
CCTCATGACATAGGCCCTCCCGACCTTTCCCTAGGTACTCAAATAGATAGGGACCTAGTT
· S T D L K A S S D D S N Y G W E D S K Y
4081 CTCTACCGACCTGAAGGCATCGTCTGACGATTCTAACTACGGCTGGGAAGACTCCAAGTA
GAGATGGCTGGACTTCCGTAGCAGACTGCTAAGATTGATGCCGACCCCTTCTGAGGTTTCA
· I P E G A T D G S A Q P R L P A S G G A
4141 TATTCCTCGAAGGCCCGCAGGATGGGCTGCCCCAGCCCGTTTGGCCCGCTAGTGGTGGTGC
ATAAGGGCTTCCGGCGTGCCTACCCAGACGGGTCCGGCCAAAACGGGCGATCACCACCAGG
· G G N P G L Y E D L F R V S V K V K N E
4201 CCGAGGAAACCCCGTCTGTACGAGGATCTTTTCCGCGTCTCTGTGAAGGTCAGAAACAC
GCCTCTCTTGGGGCCACATGCTCTAGAAAAGGCCAGAGACACTTCCAGTTCTTTGTC
· G N V A G D E V P Q L Y V S L G G P N E
4261 GGGCAAIGTCCCGGTGATGAAGTTCTCAGCTGTACGTTTCCCTAGGCGGCCCGAATGA
CCCGTTACAGCGGCCACTACTTCAAGGAGTCGACATGCAAAGGGATCCGCCGGGCTTACT
· P K V V L R K F E R I H L A P S Q S A V
4321 GCCCAAAGGTGGTACTGCGCAAGTTTGAAGCGTATTCCTTGGCCCCCTCCGAGGAGGCCGT
CGGTTTCCACCATGACCGGTTCAAATTCGCATAAGTGAACCGGGGAAGCGTCTCTCCGGCA
· W T T T L F R R D L A N W D V S A Q D W
4381 GTGGACAACGACCCCTACCCGTCGTGACCTTGCAAAACGGGACGTTTCCGGCTCAGGACTG
CACCTGTTCCTGGGAATGGGCAGCACTGGAACGTTTGACCCCTGCAAAGCCGAGTCTCTGAC
· T V T P Y P K T I Y V G N S S R K L P L
4441 GACCGTCACTCCTTACCCCAAGACGATCTACGTTGAAACTCCTCACGGAAACTGCCCGT
CTGGCAGTGAGGAATGGGGTTCTGCTAGATGCAACCTTTGAGGAGTGCCTTTGACGGCGA
· Q A S L P K A Q *
4501 CTAGGCTCGCTGCCTAAGGCCAGTAA
GGTCCGGAGCGACGATTCGGGTCATT

```

Fig. 15D

RESUMO

“MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO SECRETADO
TENDO ATIVIDADE BIOLÓGICA, PROTEÍNA DE FUSÃO ISOLADA,
POLINUCLEOTÍDEO ISOLADO, CONSTRUÇÃO DE PROTEÍNA DE
5 FUSÃO, VETOR DE EXPRESSÃO, CÉLULA HOSPEDEIRA FÚNGICA,
MÉTODOS PARA DEGRADAR OU CONVERTER UM MATERIAL
CELULÓSICO E PARA PRODUZIR UMA SUBSTÂNCIA”

A presente invenção refere-se a métodos para produzir um polipeptídeo secretado tendo atividade biológica, compreendendo: (a)
10 transformar uma célula hospedeira fúngica com uma construção de proteína de fusão codificando uma proteína de fusão, que compreende: (i) um primeiro polinucleotídeo codificando um peptídeo de sinal; (ii) um segundo polinucleotídeo codificando pelo menos um domínio catalítico de uma endoglucanase ou uma porção do mesmo; e (iii) um terceiro polinucleotídeo
15 codificando pelo menos um domínio catalítico de um polipeptídeo tendo atividade biológica; em que o peptídeo de sinal e pelo menos o domínio catalítico da endoglucanase aumenta a secreção do polipeptídeo tendo atividade biológica comparado com a ausência de pelo menos o domínio catalítico da endoglucanase; (b) cultivar a célula hospedeira fúngica
20 transformada sob condições apropriadas para produção da proteína de fusão; e (c) recuperar a proteína de fusão, um componente da mesma, ou uma combinação da mesma; tendo atividade biológica, a partir do meio de cultivo.